

Univerzita Karlova
Přírodovědecká fakulta

Studijní program: Chemie

Studijní obor: Biofyzikální chemie



Bc. Tereza Neuwirthová

Biofyzikální charakterizace proteinových knihoven z různých repertoárů aminokyselin

Biophysical characterization of protein libraries composed of different amino acid repertoires

Diplomová práce

Vedoucí závěrečné práce: Mgr. Klára Hlouchová, Ph.D.

Praha, 2020

Prohlášení

Prohlašuji, že jsem tuto závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyly předloženy k získání jiného nebo stejného akademického titulu.

V Praze dne 15.6.2020

Podpis:

Poděkování:

Nejprve bych ráda poděkovala své školitelce Mgr. Kláře Hlouchové, Ph.D. za pomoc, ochotu, trpělivost, cenné rady a podnětné prostředí v její laboratoři.

Dále bych ráda poděkovala všem kolegům z naší laboratoře, zejména Mgr. Vjačeslavu Tret'jačenkovi, za jejich neocenitelnou pomoc a rady během mé laboratorní práce.

Můj dík patří také RNDr. Lucii Bednárové, CSc. za pomoc při měření a interpretaci CD spekter.

V neposlední řadě děkuji také své rodině za velkou podporu během celé doby mého studia.

Abstrakt

Tato diplomová práce je součástí projektu zabývajícího se evolucí genetického kódu společně se studiem struktury a případné funkce prebiotických proteinů. Repertoár aminokyselin v prvotních proteinech se pravděpodobně v čase měnil a měl tak vliv na vývoj struktury a funkce bílkovin. Prvotní aminokyselinová abeceda byla zřejmě pouze poloviční velikosti dnešní aminokyselinové abecedy a pravděpodobně obsahovala pouze deset z dnešních dvaceti aminokyselin. Zmiňovaných deset aminokyselin bylo prebioticky dostupných převážně ze dvou typů potenciálních zdrojů, endogenních a exogenních.

Cílem této práce je příprava (s využitím bezbuněčné exprese) a purifikace dvou randomizovaných proteinových knihoven (obsahujících zhruba 10^{11} variant) s různým aminokyselinovým složením, a následné srovnání tendence k tvorbě sekundárních a terciárních struktur. Knihovna první totiž obsahuje pouze deset prebioticky dostupných aminokyselin, knihovna druhá všech dvacet, které dnes tvoří genetický kód.

Tento projekt by měl pomoci odhalit, jaký přínos mělo zabudování jednotlivých aminokyselin pro vznik struktury bílkovin. Celý výzkum by tak teoreticky mohl přispět k zodpovězení základních otázek nejen z oboru proteinové evoluce, ale i syntetické biologie či proteinového inženýrství.

Klíčová slova: proteinové knihovny, proteinová evoluce, struktura bílkovin, evoluce genetického kódu, syntetická biologie

Abstract

This study is part of a project which aims to understand evolution of genetic code together with structural and functional analysis of prebiotic proteins. The repertoire of amino acids in the first proteins was probably developing in time and it influenced the development of structure and function of today's proteins. First amino acid alphabet was apparently only half of the size of present alphabet, which contains twenty amino acids. These ten amino acids were probably prebiotically available from endogenous and exogenous sources.

This work includes cell-free expression and purification of two randomized protein libraries (containing approximately 10^{11} variants) with various amino acid composition and following comparison of their propensity to form secondary (using circular dichroism) and tertiary (using proteolytical analysis of sequences) structures. First library contains only ten probably prebiotically available amino acids; second library contains all twenty amino acids in today's genetic code.

This project could help us understand benefits of genetic code expansion in terms of developing structure in protein sequences. The whole research could theoretically contribute a few basic questions not only in the fields of protein evolution but also in areas of synthetic biology or protein engineering.

Key words: protein libraries, protein structure, protein evolution, genetic code evolution, synthetic biology

This thesis is written in Czech.

Obsah

Abstrakt.....	1
Abstract.....	2
Použité zkratky.....	5
1. Teoretický úvod	8
1.1. Evoluce Proteinů	8
1.2. Původ terestriálních aminokyselin	10
1.2.1. Možné endogenní zdroje aminokyselin	11
1.2.2. Možné exogenní zdroje aminokyselin	14
1.3. Aminokyseliny rané a pozdní	15
1.4. Vliv aminokyselinového složení na vlastnosti proteinů.....	21
1.5. Vlastnosti sekvenčního prostoru proteinů	22
2. Cíle práce	26
3. Experimentální metody	27
3.1. Pufrý, přístroje a chemikálie	27
3.1.1. Pufrý a ostatní roztoky	27
3.1.2. Přístroje	28
3.1.3. Chemikálie	28
3.2. Návrh a příprava knihoven.....	32
3.2.1. Návrh knihoven v DNA formě.....	32
3.2.2. Příprava dvouřetězcové DNA	33
3.2.3. Polymerázová řetězová reakce (PCR) pro přidání His-kotvy	34
3.2.4. Transkripce knihoven.....	35
3.2.5. Precipitace RNA	35
3.2.6. Translace knihoven	36
3.2.7. Afinitní purifikace knihoven s využitím FLAG nosiče.....	36
3.2.8. Denaturační purifikace knihoven s využitím TALON nosiče.....	36
3.3. Charakterizace knihoven.....	37
3.3.1. Hmotnostní analýza proteinových knihoven (technika MALDI-TOF).....	37
3.3.2. CD spektroskopie proteinových knihoven	37
3.3.3. Proteolytické štěpení	38
3.4. Návrh a příprava kontrol pro proteolytické štěpení	38
3.4.1. Polymerázová řetězová reakce (PCR).....	38

3.4.2. Restrikční štěpení genu a klonování do plasmidu pET24a	39
3.4.3. Transformace pro zmnožení plasmidové DNA.....	40
3.4.4. Izolace plasmidové DNA	40
3.4.5. Transformace konstruktů do kompetentních buněk	41
3.4.6. Exprese konstruktů a centrifugace	41
3.4.7. Sonikace	42
3.4.8. Purifikace pomocí afinitní chromatografie na TALON nosiči.....	42
3.4.9. Dialýza	42
3.4.10. Koncentrování proteinů.....	42
3.4.11. Proteolytické štěpení	42
3.5. Obecné metody	43
3.5.1. SDS elektroforéza (SDS-PAGE)	43
3.5.2. Negativní barvení SDS-PAGE gelu	44
3.5.3. Elektropřenos proteinů na nitrocelulózovou membránu s chemiluminiscenční detekcí protilátky (Western blot).....	44
3.5.3.1. Elektropřenos	44
3.5.3.2. Detekce	44
3.5.4. Horizontální agarosová elektroforéza	45
3.5.5. Akrylamidová denaturující elektroforéza	45
4. Výsledky	45
4.1. Příprava DNA a RNA knihoven	45
4.2. Translace knihoven	49
4.3. Purifikace proteinových knihoven	51
4.5. CD spektroskopie proteinových knihoven.....	54
4.6. Charakterizace tvorby terciární struktury	56
4.6.1. Kontrolní proteiny pro průkaznost specifity štěpení enzymu thrombinu.....	56
4.6.2. Optimalizace podmínek štěpení	57
4.6.3. Štěpení kombinatoriálních knihoven enzymem thrombinem.....	59
4.7. Hmotnostní analýza proteinových knihoven (technika MALDI-TOF).....	60
5. Diskuze	63
6. Závěr	67
7. Seznam literatury	68
Dodatek	73

Použité zkratky

10E	knihovna složená z aminokyselin A, D, E, G, I, L, P, S, T, V
20F	knihovna složená ze všech 20 proteinogenních aminokyselin
Ala (A)	alanin
aaRS	aminoacyl-tRNA synthetasové rodiny
AMK	aminokyseliny
Arg (R)	arginin
Asn (N)	asparagin
Asp (D)	kyselina asparagová
ATP	adenosintrifosfát
APS	peroxodisíran amonný
BSA	hovězí sérový albumin (z angl. Bovine Serum Albumin)
CD	cirkulární dichroismus
Cys (C)	cystein
DNA	deoxyribonukleová kyselina
<i>E. Coli</i>	bakterie <i>Escherichia coli</i>
EDTA	kyselina ethylendiamintetraoctová
FLAG	specifická sekvence pro afinitní purifikaci (aminokyselinová sekvence DYKDDDDK)
Gln (Q)	glutamin
Glu (E)	kyselina glutamová
Gly (G)	glycin
His (H)	histidin
HOMO	nejvyšším obsazeným molekulových orbitalem (z angl. Highest Occupied Molecular Orbital)

HPLC	vysokoúčinná kapalinová chromatografie (z angl. High-Performance Liquid Chromatography)
IDPs	neuspořádané proteiny (z angl. Intrinsically Disordered Proteins)
Ile (I)	isoleucine
LB	Luria-Bertani medium pro růst bakterií
Leu (L)	leucin
LUCA	poslední společný předek (z angl. Last Universal Common Ancestor)
LUMO	nejnižším neobsazený molekulový orbital (z angl. Lowest Unoccupied Molecular Orbital)
Lys (K)	lysin
MALDI-TOF	technika hmotnostní spektrometrie v kombinaci s detektorem doby letu (z angl. Matrix Assisted Laser Desorption/Ionization - Time of Flight)
mRNA	mediátorová ribonukleová kyselina
Met (M)	methionin
NBPs	nepřirodní randomizované proteinové sekvence (z angl. Never Born Proteins)
NEB	New England Biolabs
NMR	nukleární magnetická rezonance (z angl. Nuclear Magnetic Resonance)
NRMSD	normalizovaná odchylka
Phe (F)	fenylalanin
Pro (P)	prolin
RBS	místo pro nasednutí ribozomu v sekvencích knihoven
RNA	ribonukleová kyselina
SDS	dodecylsírán sodný
SDS-PAGE	elektroforéza v polyakrylamidovém gelu v přítomnosti dodecylsíranu sodného
Ser (S)	serin
TEMED	tetramethylethyldiamin
TFE	2,2,2-trifluoroethanol

Thr (T)	threonin
tRNA	transferová ribonukleová kyselina
Trp (W)	tryptofan
Tyr (Y)	tyrosin
UV	ultrafialové záření
Val (V)	valin

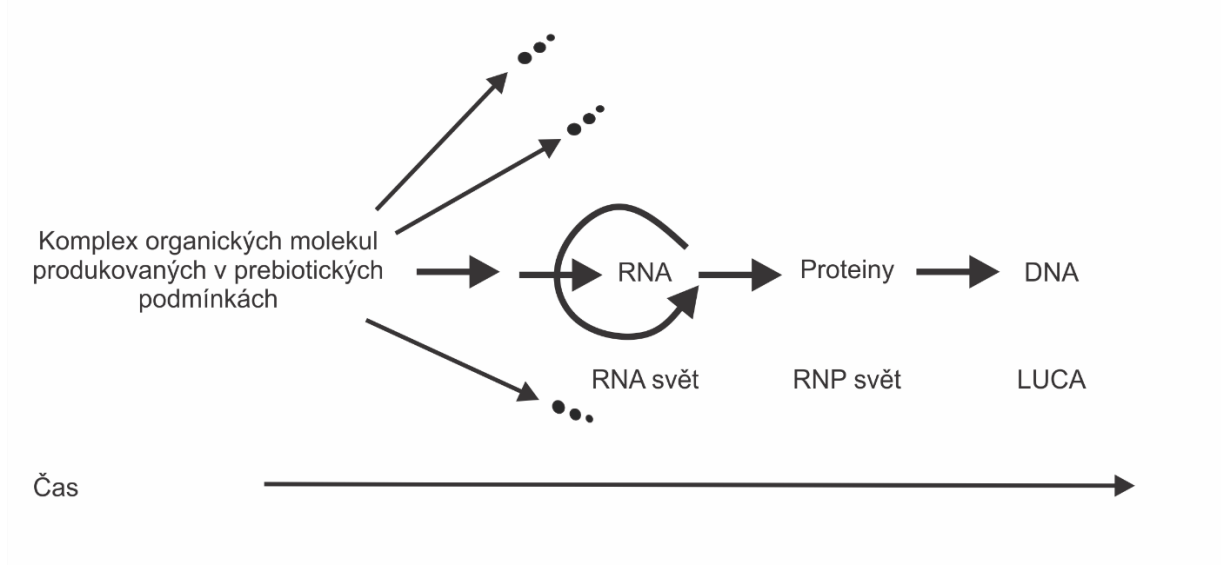
1. Teoretický úvod

1.1. Evoluce Proteinů

Během biologické evoluce se na planetě Zemi vytvořil systém, který geneticky kódovanou informaci přetváří do formy proteinů, které jsou tvořeny dvaceti různými aminokyselinami. Jednotlivé proteiny se mezi sebou liší délkou, aminokyselinovým zastoupením a sekvencí, což má přímý vliv na jejich strukturu a funkci. Aminokyselinová abeceda zůstala od posledního univerzálního společného předka (LUCA, z angl. Last Universal Common Ancestor) téměř nezměněna i přesto, že došlo k rozvoji organismů do velmi rozmanitých forem¹.

Těchto dvacet proteinogenních aminokyselin je možno kombinovat do velkého množství teoretických polypeptidových řetězců, z nichž pouze malá část se skutečně vyskytuje v živé přírodě². Z toho vyplývá stále otevřená otázka proteinové evoluce, tedy co přesně určilo výběr a následnou evoluci současných proteinů.

Prvotní hypotézy předpokládaly, že dnes existující proteiny (či jejich základní strukturní prvky) vynikaly schopností tvořit specifické třídimenzionální struktury. Proto se pouze tyto proteiny staly základními kameny života a postupem času si vyvinuly konkrétní biologickou funkci³. Obecně se totiž předpokládalo, že funkce proteinu je úzce spjata s jeho strukturou, a proto tedy pouze proteiny s definovanou strukturou mohou být biologicky funkční. Tyto teorie byly dále rozvíjeny, zejména na základě poznatků a teorií o RNA světě⁴. Schéma vývoje života v závislosti na čase, které je v souladu s touto hypotézou, je znázorněno na Obr. 1. Scénář RNA světa předpokládá dominanci RNA v rané fázi života z čehož vyplývá možnost, že prvotní peptidy byly selektovány na základě interakce a koevoluce struktury proteinů s RNA⁵.



Obr. 1 Schéma vývoje života dle teorie RNA světa. Převzato a upraveno⁴.

Rozsáhlá studie nejvíce konzervovaných strukturních motivů v souladu s touto teorií identifikovala asi 40 super-sekundárních fragmentů v sekvenci napříč existujícími proteiny, z nichž asi čtvrtina náležela k RNA/DNA vazebným úsekům (přičemž 8 z nich pocházelo z ribozomu). Tyto úseky jsou všeobecně považovány za nezbytné pro další vývoj a tím pádem také za jedny z nejstarších⁵.

Zhruba posledních 20 let výzkumu proteinů však ukázalo, že definovaná struktura nemusí nutně být prerekvizitou pro funkčnost proteinu^{6,7}.

Na základě studia vnitřně neuspořádaných proteinů (IDPs, z angl. Intrinsically disordered proteins) přichází v úvahu nový model možné evoluce prebiotických proteinů v ty dnešní. Je možné, že mezi vlastnosti, které předurčovaly raný vývoj, patřily hlavně konformační diverzita a funkční promiskuita prvotních proteinů. Tato skutečnost popírá dříve všeobecně uznávané paradigma, že jeden protein může mít pouze jednu funkci⁷. Pokud vezmeme v úvahu pouze enzymy a propojíme katalyzovanou reakci se specifickým proteinem s definovanou sekvencí a strukturou, zjistíme, že toto spojení je zřídka v poměru jedna ku jedné. Některé enzymy katalyzují více reakcí a také některé reakce využívají více různých enzymů. Enzymy s podobnostmi v sekvenci budou mít pravděpodobně stejný nebo velmi podobný mechanismus účinku. Během evoluce se dva proteiny se společným předkem (a tedy příbuznou funkcí) mohly v podobnosti v sekvenci i výrazně lišit, avšak podobnost ve struktuře stále zůstane zachována⁸.

Nový pohled na tuto problematiku nám předkládá možnost koevoluce proteinové struktury a funkce namísto teorie, že proteinová struktura předcházela vývoji funkce. Spektroskopie nukleární magnetické rezonance (NMR, z angl. Nuclear Magnetic Resonance) také prokázala, že proteiny jsou schopny v roztoku bez přítomnosti ligandu zaujmout různé konformace. Například N-koncová SH3 doména proteinu drk u rodu *Drosophila* je schopna zaujmout až šedesát různých stabilních konformací⁷.

Podrobný náhled do evoluce proteinů byl v posledních letech vnesen studiem struktury ribozomu. Jedná se o jeden z nejstarších molekulárních systémů v živých organismech, který datujeme dříve než LUCA⁹. Kvůli původu a odolnosti vůči mutacím v průběhu evoluce a díky pevně definované třídimenzionální struktuře se ribozom nabízí jako vhodný nástroj ke studiu vývoje proteinů do dnešní podoby. Ribozomální proteiny lze brát jako vhodné fosilie pro výzkum evoluce¹⁰. Z několika analýz skupiny A. Lupase a L. Williamse vyplývá, že konformace ribozomálních proteinů se liší v závislosti na vzdálenosti od peptidyl-transferázového centra, které se nachází uvnitř velké podjednotky ribozomu. Tato oblast je právě označována za nejstarší strukturní a funkční ribozomální část¹¹. Proteiny/peptidy vyskytující se v blízkosti tohoto centra jsou strukturně neuspořádané a směrem k povrchu ribozomu se nejdříve objevuje struktura skládaného listu, ke které postupně přibývá konformace helixu a tvoří se super-sekundární a sbalené struktury^{10,12,13}.

Přesný mechanismus vývoje proteinů do dnešní podoby však stále zůstává z velké části zcela neosvětlen. Rané peptidy byly zcela jistě do značné míry náhodné sekvence z dostupných aminokyselin a je nejasné, jaké vlastnosti takové peptidy mohou mít z hlediska tvorby sekundární struktury, RNA vazebného a enzymatického potenciálu. Cílem této práce je právě nahlédnutí do strukturních vlastností takových sekvencí jako první krok k zodpovězení kladených otázek. Dalším předmětem teoretického úvodu je shrnutí znalostí o původu (sekce 1.2.) a pořadí (sekce 1.3.) kódovaných aminokyselin, následované přehledem studií o vlivu aminokyselinového složení na vlastnosti proteinů (sekce 1.4.)

1.2. Původ terestriálních aminokyselin

Existuje mnoho indikátorů naznačujících, že na prebiotické planetě se vyskytovalo široké spektrum různých aminokyselin, nejen pouze těch, které se vyskytují v dnešních proteinech^{14,15}. Dosud není plně objasněno, z jakých všech možných různých zdrojů mohly tyto aminokyseliny

pocházet, pravděpodobně se jednalo o kombinaci zdrojů několika. Možné zdroje můžeme rozdělit na endogenní, které umožnily vznik aminokyselin přímo na planetě, a exogenní, díky nimž se aminokyseliny dostaly na planetu z extraterestriálního prostředí¹⁶.

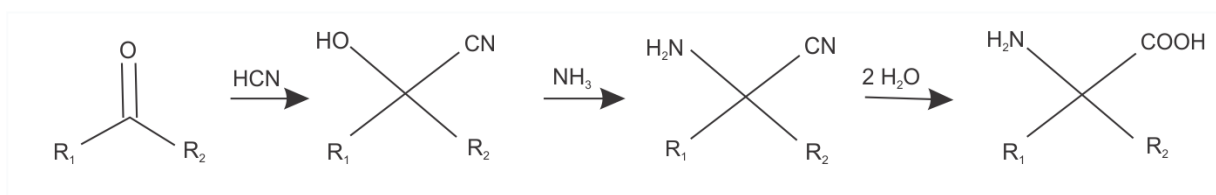
1.2.1. Možné endogenní zdroje aminokyselin

Jako jeden z možných zdrojů aminokyselin na prebiotické planetě považujeme jejich vznik přímo zde na Zemi díky působení elektrického výboje na redukující atmosféru té doby. Tato hypotéza, navržená vědci Oparinem, Millerem a Ureyem, předpokládá, že tehdejší atmosféra obsahovala převážně metan, amoniak, vodní páru a vodík. Produkty reakce iniciované zmíněným elektrickým výbojem v přítomnosti atmosférických plynů jsou kyanovodík a aldehydy, které poté mohly reagovat Streckerovou syntézou (viz Obr. 2) a formovat aminokyseliny¹⁷.

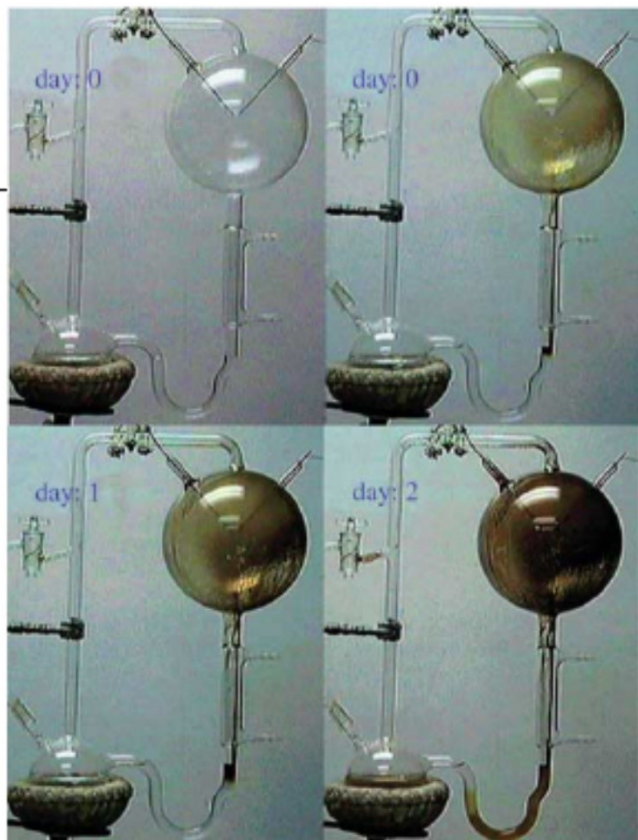
Důkaz o relevantnosti této teorie dávají Millerovy atmosférické experimenty v prostředí elektrického výboje, které slouží jako simulace možných tehdejších prebiotických podmínek. Působení elektrického výboje na směs methanu, amoniaku, vodní páry a vodíku dokazuje, že za těchto podmínek vznikají například glycin, alanin, sarkosin, β -alanin, aspartát, glutamát, laktát nebo močovina. Aparatura pro takové experimenty je znázorněna na Obr. 3. Mimo jiné tyto experimenty prokázaly, že syntéza aminokyselin za přítomnosti methanu, amoniaku a vodních par nevyžaduje konkrétní speciální podmínky. Každý proces či série procesů, které poskytly aldehydy a kyanovodík, mohly přispět k nárůstu množství aminokyselin v oceánech primitivní Země. Ukázalo se také, že stejných výsledků lze dosáhnout i použitím UV záření namísto elektrického výboje. Nezbytně nutnou podmínkou je však právě redukující atmosféra, která musela být přítomna v některém období vývoje prebiotické Země^{18,19}. V roce 2008 byly vzorky pocházející z původních Millerových experimentů znovu analyzovány pomocí modernějších technologií (HPLC (vysokoučinná kapalinová chromatografie (z angl. High-Performance Liquid Chromatography) s fluorescenčním detektorem, MALDI-TOF (technika hmotnostní spektrometrie v kombinaci s detektorem doby letu (z angl. Matrix Assisted Laser Desorption/Ionization - Time of Flight)) a bylo identifikováno ještě více aminokyselin, než se podařilo dříve Millerovi. Například valin, serin či ve velmi malém množství také fenylalanin¹⁵.

Alternativním uvažovaným atmosférickým složením v prebiotickém období planety je oxid uhelnatý, oxid uhličitý, dusík a vodní páry. energii potřebnou k iniciaci reakcí mohly dodat protony o vysoké energii pocházející z kosmického záření²⁰. Aminokyseliny mohly také vznikat

katalyzovanou syntézou při teplotách 200 až 700 °C z CO, H₂ a NH₃ v přítomnosti niklu, hliníku nebo jiných jednoduchých katalyzátorů²¹. Existuje také mnoho dalších alternativních studií zabývajících se touto problematikou, například tzv. šokovou syntézou²² či syntézou z kyanidu amonného jako jedné z výchozích látek²³.



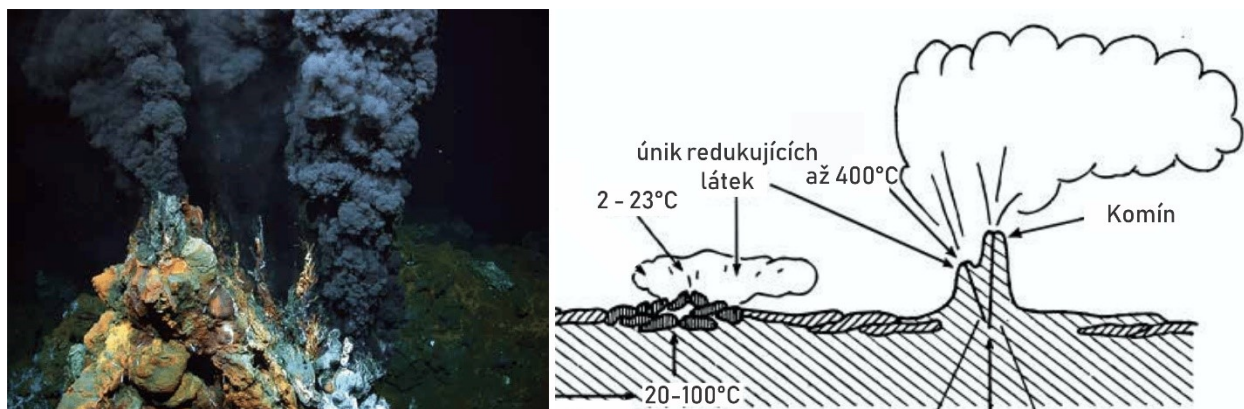
Obr. 2 Mechanismus Streckerovy syntézy. Převzato a upraveno²⁴.



Obr. 3 Příklad aparatury použité pro Millerovy atmosférické experimenty. Převzato a upraveno¹⁶.

Dalším možným endogenním zdrojem aminokyselin je možnost jejich vzniku v prostředí hydrotermálních zřidel, v jejichž okolí byly detekovány vysoké koncentrace H_2S , NH_3 , CH_4 či dokonce iontů některých kovů, což je fakt podporující tuto teorii. Znovu je pro syntézu totiž nutná přítomnost alespoň mírně redukující atmosféry. Předpokládá se také, že tyto podmořské sopky byly běžnou součástí dna dávného oceánu²⁵.

Termodynamické kalkulace provedené v roce 1998 ukazují porovnání na základě výpočtů volné Gibbsovy energie pro síť reakcí nutných ke vzniku jednotlivých aminokyselin v podmínkách v okolí hydrotermálních zřidel a v podmínkách panujících těsně pod mořskou hladinou. Mezi uvažované podmínky patří například mnohem vyšší teplota, tlak či koncentrace jednotlivých plynů. Pro 11 z 20 aminokyselin byly reakce v těchto podmínkách exergonické, v povrchové mořské vodě je endergonických všech dvacet reakcí. Mezi těchto uvedených jedenáct aminokyselin patří například leucin, valin, alanin, izoleucin nebo methionin²⁶.



Obr. 4 Fotografie a schématické znázornění hydrotermálních zřidel. Převzato a upraveno^{27,28}.

Simulace syntézy aminokyselin v podmínkách, které mohly panovat v okolí hydrotermálních zřidel byly provedeny v roce 1994 Williamem L. Marshalllem. Za použití různých reakčních směsí plynů a teplot pohybujících se okolo 200 °C se už po dvou hodinách podařilo nasimulovat vznik aminokyselin. Bylo provedeno více simulací s různými teplotami a časy, avšak všechny variace se pohybují v blízkosti výše uvedených hodnot. Ve všech simulacích došlo ke vzniku glycinu, alaninu, lysinu a izoleucinu a v převážné většině experimentů vznikl také aspartát, glutamát, prolin a serin²⁹.

Je tedy pravděpodobné, že na rané Zemi probíhala syntéza aminokyselin v různých prostředích.

1.2.2. Možné exogenní zdroje aminokyselin

Aminokyseliny pravděpodobně nepochází pouze z terestriálních zdrojů uvedených v předchozí kapitole, ale byly rovněž identifikovány v příkladných vesmírných tělesech.

Víme, že planeta Země je stará okolo 4,55 – 4,57 bilionů let³⁰. Život na Zemi pravděpodobně vznikl před 3,5 – 3,9 biliony let². Čas, který byl k dispozici pro vyvinutí života dle výše uvedených čísel tedy odpovídá 0,6 – 1 bilionu let.

Jedním ze způsobů, jak život mohl vzniknout za takto krátkou dobu je případný vznik některých základních stavebních kamenů života již v době formování planety. Chakrabarti a Chakrabarti položili otázku, zda DNA báze (konkrétně adenin) nemohou vznikat během kolapsu molekulárního mračka a formování hvězdy a pokusili se tuto hypotézu potvrdit provedením hydrodynamické simulace uvedeného procesu následovaného sledem chemických reakcí vedoucích ke vzniku prebiotických molekul. Podařilo se prokázat, že takto může vzniknout značné množství adeninu, a dokonce několik jednodušších aminokyselin, jako například glycin či alanin. Tato studie tedy naznačuje, že planety a komety už mohly vzniknout s případnou „kontaminací“ prebiotickými molekulami³¹.

Aminokyseliny byly nalezeny také na meteoritech, a to ve velmi širokém repertoáru^{14,32}. Příkladem meteoritů s vysokým zastoupením aminokyselin jsou meteority Murrayův a Murchisonův. Analýzou Murchisonova meteoritu se zabýval Michael H. Engel již v roce 1982 a i když připouští možnou terestriální kontaminaci, podařilo se mu detekovat mimo jiné kyselinu glutamovou, kyselinu asparagovou, leucin, alanin či valin³³. Dále můžeme uvést studii zabývající se analýzou meteoritu Yamato, který byl nalezen na Antarktidě, a jeho srovnáním s Murchisonovým meteoritem. Zastoupení aminokyselin se téměř neliší ani v koncentraci, ani v druzích nalezených aminokyselin³⁴. S postupem času a zlepšujícími se metodami detekce se rozšiřuje také počet aminokyselin identifikovaných na meteoritech. Zaia et al. ve své studii přehledně srovnává výskyt aminokyselin na odlišných meteoritech a z jeho studie je patrné, že na meteoritech Murchisonově, Murrayově a Yamato bylo nalezeno 11 shodných aminokyselin. Autoři studie přidali k tomuto srovnání také další tři meteority, v této práci dosud neuvedeny, a všechny obsahovaly 10 shodných aminokyselin³². Nakonec Burton et. al. ve své studii z roku 2012 uvádí více než 80 různých aminokyselin detekovaných na různých meteoritech z nichž pouze část patří mezi proteinogenní aminokyseliny¹⁴.

Zaiaova studie uvádí mimo jiné také souhrn různých typů syntézy aminokyselin za simulovaných podmínek panujících na mezihvězdném médiu jako možné exogenní aminokyselinové zdroje³². Jako příklad lze uvést studii z roku 2002, kdy se podařilo aminokyseliny detekovat i na pozůstatku analogu mezihvězdného ledu za pokojové teploty, který byl ve vakuu ozářen ultrafialovým zářením. Takto se podařilo detekovat 16 aminokyselin, například alanin, glycin, valin, prolin, serin nebo aspartát. Experimentálně se tedy potvrzuje možnost, že vesmírné prachové částice s ledovými plášti mohly být také místem syntézy aminokyselin³⁵.

Nejnovější práce, která je zatím stále v recenzním řízení, se zabývá charakterizací fragmentu identifikovaného pomocí hmotnostní spektroskopie na meteoritech Allende a Acfer086. Tato studie poukazuje na výskyt peptidového řetězce složeného z glycinových a hydroxy-glycinových molekul. Díky analýze izotopů, které tvoří tento peptidový řetězec se podle autorů podařilo vyloučit možnost, že se jedná o kontaminaci vzniklou po dopadu meteoritů na zemský povrch³⁶.

1.3. Aminokyseliny rané a pozdní

O pořadí výskytu proteinogenních aminokyselin a s tím spojené evoluci genetického kódu existuje řada různých teoretických i experimentálních studií, vycházejících z různých předpokladů, jak bylo naznačeno výše. Protože je v této vědní oblasti velmi obtížné s jistotou vyloučit či prokázat jejich podloženost a správnost, byly již počátkem 21. století provedeny rozsáhlé meta-analýzy výsledků všech těchto studií (viz následující sekce 1.3.1.). Tyto analýzy byly dále komplementovány novými přístupy (nezahrnutými v původních meta-analýzách), které jsou shrnuty v dalších podkapitolách.

Byť přesný původ a stáří aminokyselin je zpětně nemožné určit, kombinací několika přístupů bylo dosaženo v podstatě nečekaného konsenzu v rozdělení aminokyselin na rané (tedy ty, jejichž syntéza byla pravděpodobná v prebiotickém období) a pozdní (tedy ty, jejichž syntéza byla podmíněna vznikem biosyntetických drah¹).

1.3.1. Meta-analýzy experimentálních a teoretických studií původu aminokyselin

Meta-analýza zabývající se chronologií začlenění aminokyselin do genetického kódu provedená E. N. Trifonovem uvažuje celkem 40 různých kritérií v roce 2000 a dokonce 60 v roce 2004. Tato studie se neomezuje pouze na možné prebiotické zdroje aminokyselin a jejich koncentraci napříč

možnými zdroji, ale uvažuje také kritéria jako například termostabilitu raných nukleových kyselin, pak také teorie týkající se formování genetického kódu a následného propojení s proteiny, či kritéria, zabývající se analýzou rozdělení tripletů mezi aminokyselinami. Výsledné seřazení aminokyselin dle stáří je uvedeno v Tab. 1. Pořadí bylo určeno pro všech 60 kritérií, avšak některá kritéria spolu evidentně souvisí, zatímco jiná jsou na sobě kompletně nezávislá. Aby se předešlo zkreslení výsledků, související kritéria byla rozdělena do skupin a v každé skupině kritérií bylo aminokyselinám přiděleno průměrné pořadí. Dále pak byla kritéria rozdělena do dvou tříd, a to na jednotlivá nezávislá kritéria a kritéria uvažující více vzájemně závislých faktorů. Výsledné číslo získané pro každou z 20 proteinogenních aminokyselin pomocí výše naznačeného postupu je taktéž uvedeno v Tab. 1^{37,38}.

Tab. 1 Relativní stáří aminokyselin dle Trifonova³⁸ a Higgse & Pudritze³⁹. Zeleně jsou označeny aminokyseliny, které jsou v této práci (a ve většině studií) považovány za rané a žlutě aminokyseliny pozdní.

Aminokyselina	Přidělená hodnota relativního stáří (± 0.7) Trifonov	Pořadí	Přidělená hodnota relativního stáří Higgs & Pudritz	Pořadí
G	3.5	1	1.10	1
A	4.0	2	2.80	2
D	6.0	3	4.30	3
V	6.3	4	8.50	5
P	7.3	5	10.00	9
S	7.6	6	8.60	6
E	8.1	7	6.80	4
T	9.4	8	11.70	10
L	9.9	9	9.40	8
R	11.0	10	13.30	13
N	11.3	11	14.20	14
I	11.4	12	9.10	7
Q	11.4	13	14.20	14
H	13.0	14	13.30	13
K	13.3	15	12.60	11
C	13.8	16	14.20	14
F	14.2	17	13.20	12
Y	15.2	18	14.20	14
M	15.4	19	14.20	14
W	16.5	20	14.20	14

Studie Paula G. Higgse a Ralpha E. Pudritze z roku 2009 kombinuje studie zabývající se frekvencí zastoupení aminokyselin na rozličných výše uvedených zdrojích. Předestírá hypotézu, že bez

ohledu na zdroj, se zastoupení aminokyselin mezi jednotlivými zdroji nijak výrazně neliší. Deset pravděpodobně nejčastěji se vyskytujících aminokyselin napříč zdroji tato studie označuje jako rané a řadí je podle klesajícího výskytu v možných prebiotických podmínkách. Detailněji je řazení aminokyselin dle stáří ukázáno v Tab. 2. Výše uvedená studie z roku 2009 uvažuje pouze takové možné zdroje, na nichž lze přímo změřit koncentraci dané vyskytující se aminokyseliny. Sloupce M udávají zastoupení aminokyselin na meteoritech, M1 na Murchisonově meteoritu⁴⁰, M2 na Murrayově meteoritu⁴⁰ a M3 na Yamato meteoritu³⁴. Sloupec I1 udává zastoupení aminokyselin na pozůstatku analogu mezihvězdného ledu po ozáření UV zářením³⁵. Sloupce A1 a A2 reprezentují Millerovy experimenty, A1 s atmosférou složenou z methanu, amoniaku, vodní páry a vodíku, A2 s atmosférou z methanu, dusíku, vodní páry a stopového množství amoniaku⁴¹. Sloupec A3 poukazuje na diskuze o více či méně redukující atmosféře a reprezentuje právě teorie o méně redukující atmosféře spojené s iniciací reakcí pomocí vysoce energetických protonů²⁰. Vznik aminokyselin v prostředí hydrotermálních zřídél je zohledněn ve sloupcích H1 a H2. Sloupce S zahrnují další možné teorie, konkrétně Bar-Nunovu šokovou syntézu (S1)²², syntézu z kyanidu amoného (S2)⁴² a syntézu z oxidu uhelnatého, vodíku a amoniaku za přítomnosti katalyzátorů (S3)²¹. Tabulka 2 tedy ukazuje široké spektrum možných aminokyselinových zdrojů a zdůrazňuje, že i přes velmi různorodé podmínky jsou uvedená data překvapivě konzistentní. Zahrnuto je pouze 20 proteinogenních aminokyselin vyskytujících se v dnešních proteinech.

Číselná pořadí aminokyselin uvedené v Tabulce 2 ve všech sloupcích kromě H1, H2 a H3 jsou normalizována dle koncentrace glycinu. Koncentrace glycinu je vždy nejvyšší, tato aminokyselina má tedy přiřazené číslo 1. Výskyt posledních šesti aminokyselin uvedených v obou tabulkách společně v jednom řádku nepozorujeme na žádném z výše uvedených ať už endogenních či exogenních zdrojů a tyto aminokyseliny označujeme jako pozdní.

Relativní četnost aminokyselin v prebiotických podmínkách uvedené ve studiích Higgse a Pudritzeho a Trifonova pravděpodobně reflektuje aminokyselinové složení prvotních proteinů v čase kdy se formoval genetický kód. Zbylé aminokyseliny mohly být do proteinů začleněny později během vývoje dalších biochemických procesů³⁹.

Obě popsané meta-analýzy se pouze mírně liší, pokud srovnáme pořadí přiřazené konkrétním aminokyselinám, jak je patrné z Tab. 1.

Tab. 2 Relativní koncentrace AMK na možných prebiotických zdrojích. Převzato a upraveno³⁹.

Relativní koncentrace aminokyselin na možných prebiotických zdrojích													
		Meteority			Ledové částice	Atmosférické syntézy			Hydrotermální zřídla		Ostatní možné chemické syntézy		
Aminokyselina		M1	M2	M3	I1	A1	A2	A3	H1	H2	S1	S2	S3
G	Gly	1.00	1.00	1.000	1.000	1.000	1.000	1.000	18	12	1.000	1.000	40
A	Ala	0.34	0.40	0.380	0.293	0.540	1.795	0.155	15	8	0.473	0.097	20
D	Asp	0.19	0.50	0.035	0.022	0.006	0.077	0.059	10	10	0.000	0.581	30
E	Glu	0.40	0.50	0.110	0.000	0.010	0.018	0.000	6	11	0.000	0.000	20
V	Val	0.19	0.30	0.100	0.012	0.000	0.044	0.000	1	0	0.006	0.000	2
S	Ser	0.00	0.00	0.003	0.072	0.000	0.011	0.018	8	11	0.000	0.154	0
I	Ile	0.13	0.00	0.060	0.000	0.000	0.011	0.000	8	9	0.000	0.002	4
L	Leu	0.04	0.00	0.035	0.000	0.000	0.026	0.000	3	0	0.001	0.002	7
P	Pro	0.29	0.10	0.000	0.001	0.000	0.003	0.000	9	0	0.000	0.000	2
T	Thr	0.00	0.00	0.003	0.000	0.000	0.002	0.000	2	0	0.000	0.002	1
K	Lys	0.00	0.00	0.000	0.000	0.000	0.000	0.000	7	0	0.000	0.000	14
F	Phe	0.00	0.00	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	1
R	Arg	0.00	0.00	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	15
H	His	0.00	0.00	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	15
NQCVMW		0.00	0.00	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	0
Normalizace sloupců s výjimkou H1, H2 a S3 provedena pomocí předpokladu, že koncentrace glycinu, která je vždy nejvyšší, je rovna 1.00.													
Pro sloupce H1, H2 a S3 uvedené číslo značí počet experimentů při kterých byla daná aminokyselina pozorovaná ve větším než stopovém množství.													

1.3.2. Analýza současných genomů

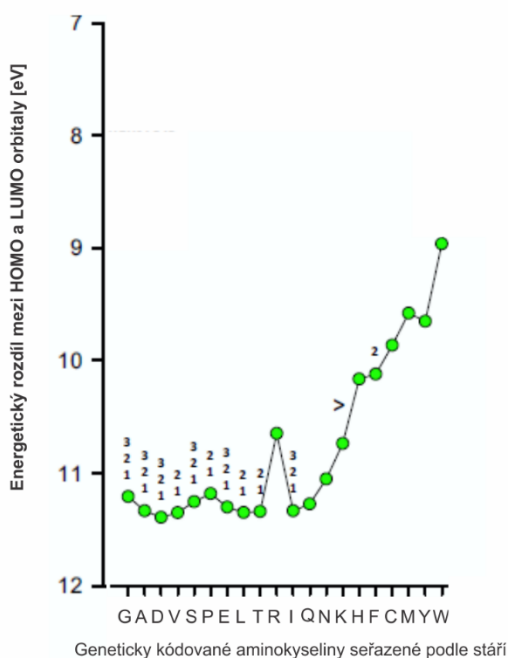
Je možné, že genetický kód nebyl v současné „kompletní“ podobě ani v čase existence LUCA. Jeden z koevolučních modelů rozrůstání genetického kódu nabízí vysvětlení, že některé aminokyseliny byly do genetického kódu začleněny během rozdělení aminoacyl-tRNA synthetasových (aaRS) rodin. Tuto teorii ve svém článku rozvádějí G. P. Fournier a E. J. Alm a srovnávají enzymy TrpRS a TyrRS, vzhledem k předpokladu, mimo jiné i Trifonova, jak je uvedeno v Tab. 1, že tryptofan v genetickém kódu relativně nová aminokyselina. TrpRS a TyrRS jsou paralogní proteinové rodiny, ve kterých zůstává zachována významná podobnost v sekvenci. Odvozené sekvenční složení jejich předchůdce nám pomáhá prokázat možnou existenci výše uvedeného koevolučního modelu, a to díky absenci tryptofanu v TyrRS, tedy pravděpodobném předchůdci TrpRS napříč organismy⁴³.

Dalším možným důkazem hovořícím ve prospěch rozšíření genetického kódu o další aminokyseliny během jeho vývoje je aminokyselinová analýza 15 setů ortologních proteinů v příbuzných genomech a identifikace rozdílného aminokyselinového složení v daných genomech, které mohla způsobit pouze změna jediné báze v tripletu a následné porovnání 75 takových aminokyselinových párů. Pokud vezmeme v úvahu všechny tyto substituce, bez ohledu na počet nahrazených nukleotidů, došlo k nárůstu pěti aminokyselin (Cys, Met, His, Ser a Phe) a ztrátě čtyř (Pro, Ala, Gly, Glu) napříč druhy. Všechny čtyři aminokyseliny, jejichž výskyt poklesl lze najít ve výsledcích analýz Murchisonova meteoritu a Millerova experimentu. Z aminokyselin, jejichž koncentrace narostla, pozorujeme ve výsledcích Millerova experimentu pouze serin a žádná z nich nebyla nalezena na Murchisonově meteoritu⁴⁴.

1.3.3. Kvantově-chemická analýza proteinogenních aminokyselin

Studie z roku 2018 nabízí analýzu kvantové chemie proteinogenních aminokyselin⁴⁵. Energetický rozdíl mezi nejvyšším obsazeným molekulovým orbitalem (HOMO, z angl. Highest Occupied Molecular Orbital) a nejnižším neobsazeným molekulovým orbitalem (LUMO, z angl. Lowest Unoccupied Molecular Orbital) je široce aplikovatelným teoretickým prediktorem chemické reaktivity pro danou sloučeninu. Čím je tento energetický rozdíl mezi HOMO a LUMO nižší, tím snadněji se molekula aktivuje a stává se reaktivnější. Tento energetický rozdíl byl změřen pro všech dvacet proteinogenních aminokyselin. Změřené hodnoty vynesené do grafu (viz Obr. 5) vykazují pozoruhodný trend, který souvisí se stářím aminokyseliny z hlediska začlenění do

genetického kódu. Aminokyseliny začleněné do genetického kódu později mají energetický rozdíl mezi HOMO a LUMO mnohem menší než aminokyseliny nalezené na Murchisonově meteoritu, či vzniklé při Millerových experimentech. Studie také potvrzuje hypotézu navrženou na základě analýz energetického rozdílu mezi HOMO a LUMO, že redoxní aktivita aminokyselin se zvyšuje s pozdější dobou začlenění aminokyseliny do genetického kódu. A po změření reaktivity vůči peroxidovým radikálům bylo potvrzeno, že pozdní aminokyseliny jsou vůči těmto radikálům reaktivnější než aminokyseliny rané. Tato studie tedy naznačuje, že genetický kód se změnil v reakci na zvýšení koncentrace kyslíku v atmosféře⁴⁵.



- 1 Aminokyseliny nalezené na meteoritech
- 2 Aminokyseliny připravené Millerovými experimenty
- 3 Aminokyseliny připravené simulacími podmínek hydrotermálních zřídél
- > Přechod mezi ranými aminokyselinami a těmi co se do genetického kódu začlenily později

Obr. 5 HOMO-LUMO energetický rozdíl proteinogenních aminokyselin. Stáří aminokyselin dle meta-analýzy provedené Trifonovem. Převzato a upraveno⁴⁵.

1.4. Vliv aminokyselinového složení na vlastnosti proteinů

1.4.1. Prvotní hypotézy

Výskyt motivů sekundární struktury v prvotních proteinech, starších než LUCA, lze jen těžko dedukovat. Z výše popsaných studií zabývajících se dostupností aminokyselin v prebiotických podmínkách (viz sekce 1.2.) si lze utvořit přibližnou představu o tom, jaké aminokyseliny převažovaly v prvotních proteinech. Dnes také víme, jak zastoupení jednotlivých aminokyselin v proteinu ovlivní jeho výslednou sekundární strukturu. Díky těmto poznatkům se lze pokusit o odhad výskytu motivů sekundárních struktur.

Předpokládáme, že pokud prebiotické proteiny byly schopny formace sekundární struktury, jednalo se spíše o β -skládaný list než o α -helix. Nejzastoupenější aminokyselinou v prebiotických podmínkách byl pravděpodobně glycin a přítomnost glycinu často narušuje strukturu α -helixu⁴⁶. Zdá se, že docházelo k formaci spíše kratších polypeptidových řetězců, což znovu ukazuje na vyšší pravděpodobnost formace β -skládaného listu⁴⁷. Tato konformace podle prací A. Brack et al. také lépe odolává teplotním výkyvům, a proto méně často podléhá chemické degradaci^{46,47}.

1.4.2. Proteiny s redukováným repertoárem aminokyselin

Z analýz výskytu aminokyselin v prebiotickém období také vyplývá, že pravděpodobně nebylo dostupných všech dvacet aminokyselin, které tvoří dnešní proteiny. Je možné, aby funkční a strukturovaný protein byl tvořen redukovanou aminokyselinovou abecedou?

Už v roce 1997 se D.S. Riddle a D. Baker pokusili redukovat aminokyselinovou abecedu na pouhých pět aminokyselin (lysin, isoleucin, glycin, alanin a kyselinu glutamovou) a těmito aminokyselinami nahradit většinu ostatních aminokyselin v malé proteinové doméně SH3 se sekundární strukturou β -skládaného listu. Poté proteiny analyzovali pomocí NMR a CD spektroskopie, díky čemuž zjistili, že i protein s redukovanou aminokyselinovou abecedou je schopen vytvořit téměř identickou strukturu jako přirozeně se vyskytující SH3 doména⁴⁸.

O charakterizaci proteinů kódovaných redukovanou aminokyselinovou abecedou se pokusili také Liam M. Longo, Jihun Lee a Michael Blaber. Použili existující protein tvořící strukturu “ β -trefoil” a vytvořili mutanty tohoto proteinu tvořené pouze ze 13 aminokyselin. Procentuální zastoupení raných aminokyselin u použitých mutantů bylo okolo 80%. Schopnost tvořit strukturu u nově vytvořených proteinů zůstala v prostředí s vysokým obsahem minerálních solí zachována. Výsledky jejich pokusu podporují hypotézu, že dostupný set aminokyselin v prebiotickém období

je dostačující k vytvoření proteinu s definovanou strukturou. Studie se také zmiňuje, že set prebiotických aminokyselin obsahuje pozoruhodně mnoho těch se sklonem k tvorbě základních typů sekundárních struktur⁴⁹. K podobným závěrům došla i nedávná rozsáhlejší výpočetní studie mnoha různých jednodoménových proteinů⁵⁰.

1.4.3. Náhodné proteinové sekvence

Ve snaze definovat obecná pravidla vlivu aminokyselinového složení (vlivem evoluce genetického kódu) na chování proteinů byly syntetizovány knihovny náhodných sekvencí, které nebyly ovlivněny procesem evoluční selekce. Tyto knihovny se ve dvou různých studiích sestávaly z (i) 5 aminokyselin (glycin, alanin, kyselinu asparagovou, kyselinu glutamovou a valin), (ii) 9-12 aminokyselin a (iii) všech 20 aminokyselin, dle pořadí z hlediska začlenění do genetického kódu, publikovaném Trifonovem jak bylo popsáno v sekci 1.3.³⁷. V obou studiích byly z každé knihovny náhodně vybrané proteiny produkovány v bakterii *E. coli*^{51,52}. Obě studie se shodovaly v tom, že nejlépe rozpustné byly náhodné proteiny z první knihovny a nejméně náhodné proteiny, které obsahovaly všech dvacet aminokyselin. Opačných výsledků bylo ale dosaženo z hlediska kvantifikace průměrné exprese. Tanaka et al. pozorovali kvantitativně větší exprese u první a druhé knihovny, zatímco Newton et al. naopak u třetí knihovny. Je pravděpodobné, že to mohlo být způsobeno odlišnými parametry testování či omezenou velikostí testovaného souboru sekvencí. Newton et al. rovněž prezentovali CD spektra vybraných proteinů, nicméně jednotlivá spektra při daném testovacím souboru nevykazovala významné rozdíly mezi jednotlivými knihovnami^{52,53}.

Aby bylo možné vyvodit dostatečně podložené závěry z podobných studií, bylo by zřejmě zapotřebí analyzovat mnohem větší množství jednotlivých sekvencí či dosáhnout biofyzikální charakterizace celých proteinových knihoven. Jen tímto způsobem je možné obsáhnout reprezentativní vzorek o sekvenčním prostoru proteinů tvořených různým repertoárem aminokyselin.

1.5. Vlastnosti sekvenčního prostoru proteinů

Výzkumu vlastností sekvenčního prostoru se vědecká komunita začala věnovat začátkem tohoto století, zejména díky vývoji metod syntetické biologie a výpočetních technologií. Většina dostupných studií se zatím věnovala sekvencím tvořených z biologického repertoáru aminokyselin. Hlavní motivací těchto studií bylo zjistit, jak moc jsou současné proteiny unikátní z hlediska struktury a funkce ve srovnání se zcela náhodnými sekvencemi.

Všechny možné proteiny, které teoreticky mohly vzniknout, avšak mezi dnešními proteiny se nevyskytují jsou označovány jako „Nikdy nenarozené proteiny” (NBPs, z angl. Never Born Proteins)⁵⁴. Knihovny NBPs byly studovány různými skupinami za účelem predikce sekundární struktury a experimentální charakterizace z hlediska výskytu funkce a sbalení proteinů.

Jedna z prvních studií zabývajících se touto problematikou se pokusila o charakterizaci knihovny obsahující 10^9 náhodných sekvencí z hlediska tvorby terciární struktury⁵⁴. Jediným nenáhodným úsekem, který mohl poskytnout zkrácení výsledků této studie je thrombinové štěpné místo (aminokyseliny P, R, G) uprostřed náhodného regionu všech sekvencí této knihovny. Tento tripeptidový substrát pro enzym thrombin poskytl kritérium pro odlišení sbalených a nesbalených proteinů, protože předpokládáme, že u sbalených proteinů bude tento region více chráněn před štěpením použitou proteázou, protože bude ukrytý uvnitř kompaktní proteinové struktury a nebude tedy pro proteázu přístupný. Z původní knihovny bylo následně náhodně vybráno několik sekvencí, které nevykazovaly téměř žádnou podobnost s existujícími proteiny a ty byly exprimovány v *E. coli* a následně štěpeny. Autoři studie pozorovali, že přes 20 % sekvencí bylo odolných vůči štěpení, a došli k závěru, že tomu tak je pravděpodobně díky tvorbě kompaktní sbalené 3D struktury⁵⁴.

Výskytem sekundární struktury v náhodném sekvenčním prostoru proteinů se doposud zabývaly především bioinformatické analýzy. První z nich využívala bioinformatický prediktor proteinové struktury Rosetta^{55,56}. Generované proteinové struktury byly analyzovány mimo jiné z hlediska jejich náboje, predikce výskytu sekundární struktury či hydrofobní složení jádra. Bylo zjištěno, že většina proteinů má tendenci tvořit kompaktní třidimenzionální strukturu a stejně tak předpokládaný výskyt sekundárních struktur byl u těchto proteinů vysoký. Struktura a povrchový náboj jsou podle autorů srovnatelné s přírodními proteiny, což naznačuje podobnou stabilitu a rozpustnost. Dále bylo pozorováno, že v porovnání s přírodními proteiny predikce vykazovaly u těchto NBPs mírně větší poměr helikální struktury⁵⁶. Nicméně paralelní studie poukázala na to, že prediktor použitý v této studii (využívaný zejména pro predikci terciární struktury biologických proteinů) je pro účely predikce NBPs zřejmě značně nepřesný⁵⁷. Druhá zde uvedená *in silico* studie na toto téma byla provedena v roce 2016⁵⁸. Na 10 000 náhodných sekvencí randomizovaných na aminokyselinovém stupni byl znovu proveden výpočet předpokladu tvorby sekundární struktury a neuspořádanosti. Použitím jediného prediktoru uspořádanosti (a několika prediktorů

neuspořádanosti) proteinů bylo zjištěno, že přirozeně se vyskytující proteiny obsahují více úseků s neuspořádanou strukturou než NBPs. Obsah sekundární struktury byl podle studie podobný jako u biologických proteinů, ovšem s preferencí β -skládaného listu. To je opačný výsledek, než publikoval ve své studii Minervini et al. navzdory téměř identickým parametrům studovaných sekvencí^{56,58}. Tento nesoulad je zřejmě daný zvolenou metodikou obou studií, kde první z nich používala pro tyto účely velmi nepřesný prediktor terciární struktury a druhá využívala pouze jeden prediktor sekundární struktury. Takový přístup může být zatížen značnou chybou, která je bez srovnání s jinými prediktory jen těžko detekovatelná.

Pozdější analýza knihovny náhodných sekvencí z naší laboratoře pracovala paralelně s pěti různými prediktory sekundární struktury a vyhodnocovala rovněž korelace mezi jednotlivými nástroji⁵⁹. Bylo zjištěno, že celkový výskyt sekundární struktury náhodných sekvencí (zhruba 45 %) je velmi podobný jako u biologických proteinů (zhruba 50 %), přičemž rozdíly ve frekvenci jednotlivých motivů jsou statisticky nevýznamné. Tyto výsledky navíc byly podpořeny experimentální charakterizací vybraných NBPs⁵⁹.

Zatím velmi málo studií se zabývalo možným výskytem funkce v náhodném sekvenčním prostoru. Velmi citovaná je raná práce z laboratoře J. Szostaka, kde byly z knihoven zcela náhodných sekvencí selektovány proteiny schopny vysokoafinitní a specifické vazby ATP⁶⁰. Nedávná studie se zabývala tolerancí a vlivem NBPs v živých organismech, konkrétně v *E. coli*. Srovnáním růstu bakteriálních kmenů s obsahem náhodných sekvencí a bez nich se zjistilo, že 25 % těchto sekvencí rychlost růstu zvyšovalo a 52 % růst inhibovalo⁶¹. Interpretace této studie ovšem vyvolala značnou odbornou diskuzi⁶². Funkční potenciál NBPs a možné vlivy zcela náhodných sekvencí na biologické systémy (související s problematikou *de novo* genů) zůstávají tedy intenzivně studovanou problematikou⁶³.

Výše uvedené studie ukazují, že některé náhodné sekvence vykazují podobné vlastnosti jako přirozeně se vyskytující proteiny. Napříč studiemi lze předpokládat strukturované a kompaktně sbalené výsledné proteiny u přibližně 5 - 20 % náhodně generovaných sekvencí^{54,64}. Všechny uvedené studie předpokládají, že schopnost tvořit sekundární a terciární strukturu je obecný znak polypeptidových řetězců. Dosud provedená práce s knihovnami, která je zde také uvedena ukazuje, že s náhodnými sekvencemi lze pracovat i s využitím dnes už běžných biochemických metod jako jejich exprese pomocí bakterie *E. coli*, ale také s využitím metod bioinformatických. U studií zabývajících se charakterizací

NBPs pomocí experimentálních metod byla vždy pro charakterizaci vybrána pouze malá část knihoven s náhodnými sekvencemi a dosud nebyla provedena experimentální charakterizace knihovny jako celku.

2. Cíle práce

- Příprava (transkripce, translace a následná purifikace) dvou proteinových kombinatoriálních knihoven s různým aminokyselinovým složením (rané a současné)
- Charakterizace sekundárních struktur proteinových knihoven pomocí cirkulárního dichroismu
- Odhad obsahu terciární struktury knihoven pomocí proteolytického štěpení a příprava kontrolních vzorků pro ověření této metodiky

3. Experimentální metody

3.1. Pufry, přístroje a chemikálie

3.1.1. Pufry a ostatní roztoky

- **44% roztok akrylamidu:** 6M akrylamid, 78mM bisakrylamid v HPLC vodě, pH neupravováno
- **Zásobní roztok pro akrylamidovou denaturující elektroforézu:** 50% močovina, 0,3M akrylamid, 4,4mM bisakrylamid v 1× TBE pufru, pH neupravováno
- **TBE pufr:** 89mM Tris-HCl (pH 7,6), 89mM kyselina boritá, 2mM EDTA
- **6×vzorkový pufr:** 350mM Tris-HCl, pH 6,8, 30% (v/v) glycerol, 350mM SDS, 4% (v/v) 2 mM 2-merkapt ethanol, 180μM bromfenolová modř
- **Elektrodový pufr:** 25mM Tris-HCl, 250mM glycin, 0,1% SDS, pH 6
- **Přenosový pufr:** 10% (v/v) methanol, 25mM Tris-HCl, 192mM glycin, 0,1 % SDS, pH 8,2
- **10×PBS pufr:** 1,37M NaCl, 27mM KCl, 100mM Na₂HPO₄, 18mM KH₂PO₄, pH neupravováno
- **PBST pufr:** 1×PBS pufr s 0,1% Tween 20
- **Roztok BSA:** 3% hovězí sérový albumin (BSA, z angl. Bovine Serum Albumin) v promývacím pufru
- **Močovinový pufr:** 8M močovina, 100mM fosforečnan sodný, 500mM chlorid sodný, 0,05% triton X-100, pH 8
- **Guanidiniový denaturační pufr:** 6M guanidinium chlorid, 100mM fosforečnan sodný, 500mM chlorid sodný, 0,05% triton X-100, pH 8
- **Imidazolový pufr:** 50mM fosforečnan sodný, 150mM imidazol, pH 7,4
- **Pufr A:** 150mM NaCl, 20mM TRIS, 0,05% triton X-100, pH 7,4
- **CD pufr:** 10mM TRIS, 5mM NaCl, 1mM TCEP, pH 7,4
- **Zásobní roztok thrombinu:** Thrombin z hovězí plazmy (100U/μl), 1×PBS pufr, pH 6.5
- **Purifikační pufr:** 300mM NaCl, 50mM kyselina glutamová, 50mM fosfát sodný, 0,05% triton X-100, pH 7,8

- **Promývací pufr:** 300mM NaCl, 50mM kyselina glutamová, 50mM fosfát sodný, 0,05% triton X-100, 3mM imidazol, pH 7,8
- **Eluční pufr:** 300mM imidazol, 300mM NaCl, 50mM kyselina glutamová, 50mM fosfát sodný, 0,05% triton X-100, pH 7,8

3.1.2. Přístroje

- Amersham Imager 600, GE Healthcare (UK)
- Aparatura na vertikální elektroforézu Bio-Rad (USA)
- Blokový termostat MB – 102, Bioer Technology (Čína)
- Centrifugy
 - Eppendorf 5415 R (Německo)
 - Biofuge Pico, Heraeus Instrument (Německo)
 - Allegra® X-15R, Beckman Coulter (USA)
 - Sorvall Evolution RC, Thermo Scientific (USA)
- CD spektrometr Chirascan, AppliedPhotophysics (UK)
- Inkubátory
 - IPP 400, Memmert (Německo)
 - myTEMP Mini, Benchmark Scientific (USA)
- Laboratorní váhy HL-400, A&D Engeneering, Inc. (USA)
- pH metr, XS Instruments (Itálie)
- PCR stroj GenePro, Bioer Technology (Čína)
- Skenovací přístroj Typhoon 9410 Variable Mode Imager, GE Healthcare (USA)
- Spektrofotometr NanoDrop 1000, Thermo Scientific (USA)
- Sonikátor Sonicator 3000, Misonix (USA)
- UV transiluminátor QUANTUM ST4-1100, Vilber Lourmat (Německo)
- Zdroj napětí PowerPac™ Basic power supply, Bio-Rad (USA)
- Hmotnostní spektrometr Ultraflexxtreme (Bruker, Německo)

3.1.3. Chemikálie

Bio-Rad (USA)

- Marker Precision Plus Protein Dual Color Standard

- Gel Mini-PROTEAN Tris-Tricine (16,5 %)
- Gel Mini-PROTEAN TGX (8-16 %)
- Membrána Trans-Blot Turbo Transfer Pack

Biotium, Hayward (USA)

- Fluorescenční barvivo GelRed

BioSynth AG (Švýcarsko)

- IPTG

Carl Roth (Německo)

- Arginin-HCl
- Imidazol

Fluka (Německo)

- 2-merkaptoethanol
- Akrylamid
- Dodecylsírán sodný
- TEMED
- Triton X-100

GeneFrontier (Japonsko)

- PURE*flex* 2.0

Lach-Ner (USA)

- Glycin

Lucigen (USA)

- AmpliScribe™ T7-Flash™ Transcription Kit
- OverExpress™ C41(DE3) *E. Coli* kompetentní buňky

MP Biomedicals (USA)

- Tricin

New England Biolabs (NEB, USA)

- DNA Polymerase I, Large (Klenow) Fragment
- CutSmart buffer (10X)
- 100 bp DNA ladder
- Instant Sticky-end Ligase Master Mix
- Restrikční enzym NdeI
- Restrikční enzym XhoI
- Směsný roztok deoxynukleotidů (dNTPs)
- 2×RNA loading Dye
- Low Range ssRNA marker
- NEBuffer 2.1 (10X)

Penta (Česká republika)

- Agarosa
- Boritá kyselina
- Dihydrogenfosforečnan draselný
- Ethanol
- Ethylendiamintetraoctová kyselina
- Chlorid draselný
- Chlorid sodný
- Methanol
- Močovina

Qiagen (Německo)

- GelPilot 100 bp Plus Ladder
- 5×GelPilot Loading Dye
- QIAprep Spin Miniprep Kit

Serva (Německo)

- Bromfenolová modř

Sigma-Aldrich, Německo

- 3X FLAG® Peptide
- 2,2,2-trifluoroethanol
- Acetonitril
- APS
- Anti-FLAG® M2 Magnetic Beads
- Anti-FLAG protilátka
- Bisakrylamid
- BSA
- Fosfát sodný
- Glutamová kyselina
- Glycerol
- Hydrogen fosforečnan sodný dihydrát
- Kanamycin sulfát
- LB agar
- LB médium
- Octan amonný
- Peroxodisíran draselný
- Monoclonal Anti-6X His tag antibody produced in mouse protilátka
- TCEP
- TEMED
- Thrombin from bovine plasma (40-300 NIH units/mg protein)
- Trizma báze (TRIS)
- Tris-HCl
- Tween 20

Takara Bio (USA)

- TALON afinitní nosič

ThermoFisher Scientific (USA)

- Nanášecí pufr pro agarosovou elektroforézu: 6×DNA Loading Dye
- Platinum™ PCR SuperMix High Fidelity
- One Shot™ BL21 Star™ (DE3) chemicky kompetentní *E. coli*
- One Shot™ TOP10 chemicky kompetentní *E. coli*
- Novex™ (10-20 %) Tricine, 1.0 mm, Mini Protein Gel
- SuperSignal West Femto chemiluminiscent solution

Zymo Research (USA)

- DNA (PCR) Clean-up & Concentrator kit

3.2. Návrh a příprava knihoven

3.2.1. Návrh knihoven v DNA formě

Proteinové knihovny charakterizované v této práci byly navrženy ve formě jednořetězcových DNA oligonukletidů a komerčně syntetizovány firmou Integrated DNA Technologies (IDT, USA). Tyto sekvence obsahují na svém začátku T7 promotor pro iniciaci transkripce a místo pro nasednutí ribozomu (RBS). Randomizované regiony sekvencí v knihovnách byly navrženy z degenerovaných kodonů pomocí algoritmu CoLiDe⁶⁵, který byl vyvinutý v laboratoři, kde byla tato práce vypracována. Výstupem algoritmu je přesné řešení pro kombinatoriální knihovny o zadané délce a aminokyselinovém složení na úrovni DNA templátu pomocí degenerovaných kodonů. Tímto postupem bylo dosaženo variability v knihovnách přibližně 10^{11} . Aminokyselinové složení těchto regionů v jednotlivých knihovnách a kritérium výběru zastoupených aminokyselin je uvedeno v Tab. 3 i s příslušným označením jednotlivých knihoven, které je dále používáno v této práci. Tyto regiony jsou v návrhu unikátní pro každou sekvenci v knihovně, vždy však začínají iniciačním kodonem ATG a uprostřed obsahují definovanou sekvenci sloužící ke spojení dvou sekvencí tak, aby bylo možno vytvořit dvouřetězcovou DNA (viz sekce 3.2.2.). Místo překryvu je sekvence, která je rozpoznávána a následně štěpena enzymem thrombinem a v knihovnách je také kvůli charakterizaci míry tvorby terciární struktury (viz sekce 3.3.3). Na konci jednotlivých sekvencí je FLAG kotva (aminokyselinová sekvence DYKDDDDK), která slouží k purifikaci a vizualizaci proteinů pomocí techniky Western blot (viz sekce 3.5.3). Délka

jednotlivých sekvencí v knihovnách po spojení obou oligonukletidů je 374 bází, přičemž celková délka randomizovaného úseku (i s definovanou sekvencí uprostřed) je 277 bází. Sekvence všech oligonukleotidů použitých pro syntézu jsou uvedeny v dodatku této práce.

Tab. 3 Popis jednotlivých knihoven

Označení	Kritérium výběru aminokyselin	Aminokyseliny
20F	Všech dvacet proteinogenních aminokyselin (z angl. Full)	G, A, D, V, P, S, E, T, L, R, N, I, Q, H, K, C, F, Y, M, W
10E	Deset raných aminokyselin (z angl. Early)	A, D, E, G, I, L, P, S, T, V

3.2.2. Příprava dvouřetězcové DNA

Jednotlivé navržené oligonukleotidy popsané výše (viz sekce 2.1.) byly podrobeny níže popsané reakci, aby výsledné knihovny tvořila dvouřetězcová DNA. Oligonukleotidy byly nejprve spojeny v místě překryvu bez účasti polymerázy přes teplotní gradient od 92 °C do 25 °C. Teplota se snižovala vždy po 5 minutách skokem o 2 °C. Posledním krokem bylo chlazení po dobu 5 minut při 4 °C. Složení reakční směsi je uvedeno v Tab. 4. Následně byly ke směsi přidány 2 µl polymerázy DNA Polymerase I, Large (Klenow) Fragment (NEB, USA, 5000 units/ml) a reakce probíhala dle Tab. 5, za účelem syntézy chybějící DNA na obou řetězcích od místa překryvu a vytvoření dvouřetězcové DNA knihovny. Celkový objem reakce i s přidanou polymerázou odpovídá 50 µl.

Tab. 4 Reakční směs

Složka	Objem
Vláknko 1 (100µM)	1 µl
Vláknko 2 (100µM)	1 µl
dNTPs (10mM)	1 µl
NEBuffer 2.1 (10×)	5 µl
voda	40 µl
Celkový objem reakce	48 µl

Tab. 5 Teplotní a časové fáze reakce po přidání polymerázy

Fáze	Teplota [°C]	Čas [min]
1	25	5
2	37	60
3	50	15
4	4	5

Nakonec byla výsledná DNA přečištěna kitem DNA (PCR) Clean-up & Concentrator (Zymo Research, USA).

3.2.3. Polymerázová řetězová reakce (PCR) pro přidání His-kotvy

Výsledná DNA knihovna ve formě dvouřetězcových molekul (příprava viz sekce 3.2.2.) byla použita jako templát do PCR reakce. Tato reakce byla s knihovnou provedena kvůli přidání His-kotvy na 3' konec, která slouží k denaturační purifikaci výsledných, dále charakterizovaných proteinů. Složení reakční směsi je uvedeno v Tab. 6, průběh fází PCR reakce v Tab. 7.

Tab. 6 Reakční směs

Složka	Objem
PCR SuperMix High Fidelity	45 µl
Primer A (10µM)	2 µl
Primer B (10µM)	2 µl
DNA knihovna	x µl (= 50 ng)
voda	y µl
Celkový objem reakce	50 µl

Tab. 7 Průběh fází PCR reakce

Fáze	Teplota [°C]	Čas [s]	Počet cyklů
1	55	30	1
2	95	10	24
3	55	30	
4	68	20	
5	68	120	1

3.2.4. Transkripce knihoven

Transkripce byla prováděna pomocí Ampliscribe T7 Flash Transcription kitu (výrobce Lucigen, USA), složení reakční směsi je uvedeno v Tab. 8. Reakce byla vždy namíchána při teplotě 23 °C ve stejném pořadí jako je v příslušné tabulce uvedeno. Celkový objem jedné reakce byl 20 µl. Reakční směs byla inkubována 2 h při 37 °C.

Tab. 8 Složení reakční směsi při transkripci pomocí kitu

Složka	Objem
DNA templát	x µl (= 1 µg)
voda	y µl
10×koncentrovaný pufr (Ampliscribe T7 Flash Transcription kit)	2 µl
NTPs (100 mM)	7,2 µl
DTT (100 mM)	2 µl
RNAzový inhibitor (Ampliscribe T7 Flash Transcription kit)	0,5 µl
T7 Flash roztok enzymů	2 µl
Celkový objem reakce	20 µl

3.2.5. Precipitace RNA

Ke standardní reakci pomocí Ampliscribe T7 Flash Transcription kitu (Lucigen, USA) o objemu 20 µl byl přidán tentýž objem 5M octanu amonného. Poté byla směs inkubována na ledu po dobu 20 minut. Po ukončení inkubace byla směs centrifugována (10000×g) při 4 °C po dobu 15 minut. Vzniklá peleta byla promyta 70% ethanolem a směs byla znovu centrifugována 15 minut za výše uvedených podmínek. Následně byl odstraněn téměř veškerý ethanol, zbytkové množství ethanolu ze zkumavky se nechalo odpařit a RNA byla rozpuštěna ve sterilní deionizované vodě. Koncentrace RNA po precipitaci byla měřena spektrofotometricky.

3.2.6. Translace knihoven

K translaci knihoven do formy proteinu byl v této práci využit komerční kit pro bezbuněčnou expresi PUREfrex 2.0 (GeneFrontier, Japonsko). Celkový objem jedné reakce odpovídal 20 μ l. Složení reakční směsi je uvedeno v Tab. 9. Jednotlivé roztoky potřebné pro reakci byly skladovány při – 80 °C a byly vždy rozmrazeny na ledu. Příprava samotné reakce probíhala také na ledu. Reakční směs byla inkubována při 30 °C po dobu 4 hodin.

Tab. 9 Složení reakční směsi pro translaci

Složka	Objem
Roztok I	10 μ l
Roztok II	1 μ l
Roztok III	2 μ l
RNA templát (koncentrace se různila dle použité knihovny, výsledné množství v reakci 3 μ g)	x μ l
Triton (5%)	0,2 μ l
voda	y μ l
Celkový objem reakce	20 μl

3.2.7. Afinitní purifikace knihoven s využitím FLAG nosiče

Anti-FLAG magnetický nosič (Sigma-Aldrich, Německo) byl před vnesením do reakce třikrát promyt 150 μ l pufrem A (složení viz sekce 3.1.1.). Poté byl odpovídající objem nosiče (3 μ l/reakce (20 μ l)) vnesen do traslační reakce a takto inkubován po celou dobu jejího trvání.

Poté byl magnetický nosič s navázanou knihovnou třikrát promyt 150 μ l pufru A, doba promývání byla vždy 5 minut. Následně byl protein eluován do 50 μ l pufru A pomocí přidání 2 μ l FLAG peptidu (koncentrace 5 mg/ml, Sigma-Aldrich, Německo), eluce probíhala 1 h při teplotě 23 °C za stálého míchání. Uvedené objemy odpovídají jedné translační reakci a v případě větší produkce byly zvýšeny vynásobením počtem použitých reakcí.

3.2.8. Denaturační purifikace knihoven s využitím TALON nosiče

Příslušné množství afinitního nosiče TALON (4 μ l na jednu reakci (20 μ l), Takara Bio, USA) bylo nejprve třikrát promyto guanidiniovým denaturačním pufrem (složení viz sekce 3.1.1.), poté byla translační reakce tímto pufrem desetkrát naředěna a přidána k afinitnímu nosiči. Tato směs byla za stálého míchání inkubována při teplotě 23 °C po dobu 12 hodin. Poté byl nosič dvakrát promyt močovinovým pufrem (složení viz sekce 3.1.1.) a poté dvakrát deionizovanou vodou s obsahem Tritonu X-100 0,05 %. Nakonec byla provedena eluce knihoven z nosiče do 12 μ l imidazolového

pufru za stálého míchání po dobu třiceti minut při teplotě 23 °C. Uvedené objemy odpovídají jedné translační reakci a v případě větší produkce byly zvýšeny vynásobením počtem použitých reakcí.

3.3. Charakterizace knihoven

3.3.1. Hmotnostní analýza proteinových knihoven (technika MALDI-TOF)

Knihovny připravené translační reakcí (viz sekce 3.2.6.) a purifikované pomocí denaturační purifikace na TALON nosiči (viz sekce 3.2.8.) byly podrobeny hmotnostní analýze technikou MALDI-TOF, což je technika hmotnostní spektrometrie v kombinaci s detektorem doby letu (z angl. Matrix Assisted Laser Desorption/Ionization - Time Of Flight). Přístroj byl kalibrován pomocí směsi proteinů v rozsahu 5-20 kDa. Na MALDI terčík byl nanesen 1 µl roztoku matrice kyseliny sinapové. Po zaschnutí byl na terčík nanesen 1 µl vzorku rozpuštěného v 50% acetonitrilu. Po zaschnutí byla deska s terčíkem vložena do MALDI-TOF přístroje (Ultraflextreme, Bruker, Německo). Spektra z měření byla snímána s frekvencí 1 kHz a průměrována z několika různých míst na terčíku. Výsledné spektrum bylo automaticky zpracováno pomocí programu Flexanalysis (Bruker, Německo).

3.3.2. CD spektroskopie proteinových knihoven

Vzorky pro účely měření CD spekter byly z DNA knihoven připraveny pomocí transkripce (viz sekce 3.2.4. a 3.2.5.) a translace (viz sekce 3.2.6.). Následně byly purifikovány pomocí FLAG afinitního nosiče (viz sekce 3.2.7.). Pro přípravu vzorků pro CD spektroskopii byl použit eluční pufr s takovým složením, které umožnilo zachovat vzorek a zároveň minimalizovalo absorpci pufru (CD pufr, složení viz sekce 3.1.1.) Pro charakterizaci schopnosti tvorby α -helikální konformace byly vzorky knihoven titrovány pomocí 2,2,2-trifluoroethanolu (TFE) v intervalu 0% – 50% (v/v), používaného jako rozpouštědla, schopného preferenčně tvořit helikální strukturu proteinů a peptidů⁶⁶. CD spektra byla měřena při pokojové teplotě (23 °C) za použití stolního CD spektrometru Chirascan (AppliedPhotophysics, UK). Spektra byla měřena ve spektrálním intervalu 190 nm - 300 nm s následujícím nastavením: krok 0,5 nm, rychlost monochromátoru 10 nm/min, časová konstanta 8 s, počet cyklů 3. Koncentrace knihoven byla určena pomocí aminokyselinové analýzy vzorků (pro nejvíce zastoupené aminokyseliny). Numerická analýza CD dat byla provedena pomocí programu CDPro^{67,68}.

3.3.3. Proteolytické štěpení

Knihovny připravené translační reakcí (viz sekce 3.2.6.) a purifikované pomocí afinitní purifikace na FLAG nosiči (viz sekce 3.2.7.) byly následně štěpeny pomocí enzymu thrombin (Sigma-Aldrich, Německo). Množství použitého enzymu procházelo optimalizací a je uvedeno vždy u jednotlivých reakcí ve výsledkové sekci (viz sekce 4.6.3.). Na jednu proteolytickou reakci (10 μ l eluční frakce) byla vždy použita 1/5 translační reakce (objem jedné reakce 20 μ l) po afinitní purifikaci na FLAG nosiči (celá eluční frakce 50 μ l, viz sekce 3.2.7.). K tomuto objemu byl přidán příslušný objem enzymu. Jednotlivé štěpné reakce probíhaly vždy po dobu 30 minut při 37 °C.

3.4. Návrh a příprava kontrol pro proteolytické štěpení

Sekvence proteinu s označením 665U byla již připravena a vložena do plasmidu pET24a díky předchozí práci ve skupině⁶⁹. Sekvence proteinu s označením 665C byla objednána jako syntetický gen (GeneArt High-Q String, ThermoFisher Scientific, USA). Sekvence obou proteinů je uvedena v dodatku této práce.

3.4.1. Polymerázová řetězová reakce (PCR)

PCR reakce byla použita k amplifikaci konstruktů 665C.

Postup PCR:

Do PCR mikrozkuřavky byla pipetována reakční směs o celkovém objemu 50 μ l. Složení reakční směsi je uvedeno v Tab. 10, průběh PCR reakce v Tab. 11. Sekvence použitého Primeru C a Primeru D jsou uvedeny v dodatku této práce.

Tab. 10 Složení reakční směsi pro PCR reakci

Složka	Objem
PCR SuperMix High Fidelity	45 μ l
Primer C (10 μ M)	2 μ l
Primer D (10 μ M)	2 μ l
DNA konstrukt	x μ l (= 50 ng)
voda	y μ l
Celkový objem reakce	50 μl

Tab. 11 Průběh fází PCR reakce

Fáze	Teplota [°C]	Čas [s]	Počet cyklů
1	94	30	1
2	94	15	30
3	45	20	
4	68	30	
5	68	300	1
6	4	300	1

Po PCR reakci byla DNA přečištěna pomocí kitu DNA (PCR) Clean-up & Concentrator (Zymo Research, USA).

3.4.2. Restrikční štěpení genu a klonování do plasmidu pET24a

Konstrukt pro přípravu proteinu 665C byl vnesen do plasmidu pET24a mezi místa pro restrikční endonukleázy NdeI a XhoI. Pro štěpení DNA konstruktů a plasmidu tedy byly použity restrikční endonukleázy NdeI a XhoI a pro ligaci Instant Sticky-end Ligase Master Mix (NEB, USA). Složení reakční směsi pro restrikční štěpení je uvedeno v Tab. 12, reakce byla inkubována 1 h při teplotě 37 °C. Složení reakční směsi pro ligaci je uvedeno v Tab. 13, reakce byla promíchána pipetou a inkubována po dobu pěti minut na ledu.

Tab. 12 Složení reakční směsi pro restrikční štěpení

Složka	Objem
DNA (665C protein nebo plasmid pET24a)	x μ l (= 1 μ g)
10X Cut Smart pufr (NEB)	5 μ l
enzym NdeI	1 μ l
enzym XhoI	1 μ l
voda	y μ l
Celkový objem reakce	50 μl

Tab. 13 Složení reakční směsi pro ligaci

Složka	Objem
DNA plasmid pET24a	x μ l (= 60 ng)
DNA 665C konstrukt	y μ l (= 10 ng)
voda	z μ l
Ligase Master Mix	5 μ l
Celkový objem reakce	10 μl

3.4.3. Transformace pro zmnožení plasmidové DNA

Do kompetentních buněk *E. coli* kmen TOP 10 (ThermoFisher Scientific, USA) byl vložen plasmid pET24a obsahující vložený gen za účelem zmnožení plasmidové DNA.

Ke 40 μ l kompetentních buněk byl přidán 1 μ l produktu ligace připraveného pomocí postupu popsaného v sekci 3.4.2. a následně byly buňky ponechány na ledu po dobu 30 minut. Poté byl proveden teplotní šok, během kterého byly buňky inkubovány při 42 °C v blokovém termostatu po dobu 30 s a pak znovu přesunuty na led. Na ledu byly ponechány po dobu 2 minut. Po provedení teplotního šoku bylo k buňkám přidáno 950 μ l LB média (Sigma-Aldrich, Německo). Buněčná suspence byla inkubována při 37 °C po dobu 45 minut v blokovém termostatu za neustálého třepání (200 rpm). Po uplynutí 45 minut byly buňky s LB médiem přesunuty na sterilní Petriho misku s pevným LB agarem (Sigma-Aldrich, Německo) s obsahem antibiotika kanamycinu (výsledná koncentrace 40 mg/ml). Petriho miska byla přes noc ponechána v inkubátoru při teplotě 37 °C. Následně byla založena prekultura z jedné z narostlých kolonií do 5 ml tekutého LB média s kanamycinem o stejné koncentraci. Tyto prekultury byly inkubovány v třepačce o teplotě 37 °C a 220 rpm přes noc.

3.4.4. Izolace plasmidové DNA

Pro izolaci plasmidové DNA z 5 ml kultur bakterií *E. coli* byla použita komerčně dodávaná sada QIAprep Spin (Qiagen, Německo). Bylo postupováno podle instrukcí výrobce a plasmidová DNA byla v posledním kroku eluována do 50 μ l vody. Hodnota výsledné koncentrace izolované DNA byla změřena spektrofotometricky. Délka sekvence byla ověřena pomocí metody horizontální agarosové elektroforézy (viz sekce 3.5.3.).

3.4.5. Transformace konstruktů do kompetentních buněk

Pro expresi rekombinantních proteinů byly použity kompetentní buňky *E. coli* kmen BL21 (DE3, ThermoFisher Scientific, USA), které byly skladovány při teplotě $-80\text{ }^{\circ}\text{C}$. Před samotnou transformací byly buňky vždy rozmrazeny v ledové lázni. K $40\text{ }\mu\text{l}$ kompetentních buněk byl přidán $1\text{ }\mu\text{l}$ plasmidové DNA (50 ng) obsahující požadovaný konstrukt. Pro protein 665C byl použit plasmid získaný výše popsáním způsobem a pro protein 665U byl použit plasmid již připravený díky předchozímu výzkumu skupiny, jejíž členkou je autorka práce⁶⁹. Buňky byly ponechány na ledu po dobu 30 minut. Poté byl proveden teplotní šok, během kterého byly buňky zahřáty na $42\text{ }^{\circ}\text{C}$ v blokovém termostatu po dobu 30 s a pak byly znovu přesunuty na led a zde ponechány po dobu 2 minut. Po provedení teplotního šoku bylo k buňkám přidáno $950\text{ }\mu\text{l}$ LB media. Buňky s přidáním LB médiem byly inkubovány při $37\text{ }^{\circ}\text{C}$ po dobu 1 h v blokovém termostatu za neustálého třepání (200 rpm). Po uplynutí 1 h byly buňky s LB médiem přesunuty na sterilní Petriho misku s pevným LB agarem (viz sekce 3.1.1.) s obsahem antibiotika kanamycinu (výsledná koncentrace 40 mg/ml). Petriho miska byla přes noc ponechána v inkubátoru při teplotě $37\text{ }^{\circ}\text{C}$. Následně byla založena prekultura z narostlých kolonií do 5 ml tekutého LB média s kanamycinem o výsledné koncentraci znovu 40 mg/ml . Tyto prekultury byly inkubovány v třepačce o teplotě $37\text{ }^{\circ}\text{C}$ a 220 rpm přes noc.

3.4.6. Exprese konstruktů a centrifugace

Bakteriální prekultura připravená při transformaci konstruktů byla následující den po kultivaci přelita do skleněných nádob obsahujících 500 ml tekutého LB média s kanamycinem o výsledné koncentraci 40 mg/ml . Takto připravené roztoky pro kultivaci bakterií byly umístěny do třepačky, kde při teplotě $37\text{ }^{\circ}\text{C}$ a otáčkách 220 rpm probíhal růst buněčné kultury. Průběžně byla během růstu měřena optická densita (OD) při 600 nm . Když bylo dosaženo hodnoty OD v intervalu $0,6 - 0,8$ bylo do roztoku přidáno $500\text{ }\mu\text{l}$ 1 M IPTG, aby bylo dosaženo výsledné koncentrace 1 mM . Přidáním IPTG došlo ke spuštění syntézy požadovaného proteinu, která probíhala po dobu 4 hodin v třepačce při $37\text{ }^{\circ}\text{C}$ a otáčkách 220 rpm . Po uplynutí této doby byly kultury přelity do plastových kyvet a odstředěny za teploty $4\text{ }^{\circ}\text{C}$ a $3000\times g$ po dobu 20 minut. Následně bylo odstraněno LB médium, jelikož buňky se usadily na dně kyvet a kyvety s usazenými buňkami byly umístěny do $-20\text{ }^{\circ}\text{C}$ a zde skladovány.

3.4.7. Sonikace

Proces sonikace začal rozmrazením buněk připravených metodou popsanou v sekci 3.4.6. a jejich rozpuštěním v 30 ml purifikačního pufru (viz sekce 3.1.1.). Samotná sonikace probíhala v ledové lázni (5 × 30 s při výkonu 55 W). Po sonikaci byly vzorky obsahující připravovaný protein vloženy do centrifugy a při 4 °C byly vystaveny 20 000×g po dobu 30 minut. Díky tomuto výše popsanému procesu byl získán vzorek obsahující žádaný protein bez obsahu nerozpustných částí lyzovaných buněk, které se po centrifugaci usadí na dně zkumavky.

3.4.8. Purifikace pomocí afinitní chromatografie na TALON nosiči

Oba proteiny (665C, 665U) purifikované pomocí této metody na svém N-konci nesly His-kotvu (angl. His-tag, šest histidinů za sebou) umožňující purifikaci touto metodou. Kolona určená k afinitní chromatografii byla naplněna 2 ml nosiče TALON (Takara Bio, USA) a následně třikrát promyta 10 ml purifikačního pufru (viz sekce 3.1.1.). Poté byla kolona naplněna vzorkem získaným po sonikaci a následné centrifugaci (viz sekce 3.4.7.). Po prokapání vzorku kolonou byla kolona s nosičem třikrát promyta 10 ml promývacího pufru (viz sekce 3.1.1.) a poté byl protein eluován nejprve 3 ml a následně 5 ml elučního pufru (viz sekce 3.1.1.).

3.4.9. Dialýza

Do dialyzační membrány byly přeneseny obě eluční frakce získané afinitní chromatografií. Membrána byla zajištěna svorkami a ponořena do kádinky s 500 ml PBS pufru (viz sekce 3.1.1.). Dialýza probíhala přes noc za neustálého míchání při teplotě 4 °C.

3.4.10. Koncentrování proteinů

Ke koncentrování byl použit komerční koncentrátor s póry o velikosti 10 kDa. Koncentrátor s proteinem byl centrifugován při 4 °C, 2 000×g, s intervalem 15 min. Objem byl snížen přibližně na 3 ml. Poté byla provedena kontrola čistoty proteinu pomocí SDS-PAGE (elektroforéza v polyakrylamidovém gelu v přítomnosti dodecylsírany sodného) a koncentrace změřena spektrofotometricky. Protein byl přesunut do - 80 °C, kde byl skladován pro další použití.

3.4.11. Proteolytické štěpení

Proteiny 665U a 665C byly štěpěny za použití různých koncentrací enzymu a lišila se také doba štěpení (viz sekce 4.2.2.). Každá reakce vždy obsahovala 2 µg proteinu. Následně byl přidán PBS pufr a požadované množství enzymu. Výsledný objem reakce odpovídal vždy 10 µl. Všechny reakce byly inkubovány při teplotě 37 °C. Po uplynutí požadované doby byly reakce ukončeny

zchlazením na pokojovou teplotu a přidáním vzorkového pufru (viz sekce 3.1.1.). Nakonec byly vzorky analyzovány pomocí SDS elektroforézy a Western blotu (viz sekce 3.5.2. a 3.5.3.).

3.5. Obecné metody

3.5.1. SDS elektroforéza (SDS-PAGE)

Nejprve byla sestavena aparatura a poté byly připraveny gely, jejichž složení je uvedeno v Tab. 14. Kromě gelů připravovaných v laboratoři s níže popsáním složením byly použity také komerční gradientové Tris-Glycinové (8-16 %, Bio-rad, USA) nebo tricinové gely (10-20 %, ThermoFisher Scientific, USA).

Tab. 14 Složení jednotlivých gelů pro SDS elektroforézu

Separační gel (18%)		Zaostřovací gel (6,6%)	
Složka	Objem	Složka	Objem
1,5M Tris, pH 8,8	2,5 ml	1M Tris, pH 6,8	1,25 ml
44% roztok akrylamidu	4,0 ml	44% roztok akrylamidu	750 µl
10% SDS	100 µl	10% SDS	50 µl
10% APS	100 µl	10% APS	50 µl
TEMED	10 µl	TEMED	10 µl
doplněno destilovanou vodou do 10 ml		doplněno destilovanou vodou do 5 ml	

Složky 10% APS a TEMED byly vždy přidány až těsně před nalitím gelu mezi dvě skla. Separační gel byl pipetován zhruba do tří čtvrtin výšky skel a poté byl převrstven ethanolem. Po 30 minutách stání gelu při teplotě 23 °C byl ethanol odlit a byl napipetován zaostřovací gel na již ztuhlý separační gel. Okamžitě byl vsunut hřeben pro vytváření jamek potřebných k nanášení vzorků. Po dalších 30 minutách polymerace gelu byl hřeben vyjmut a skla byla umístěna do elektroforetické aparatury a přelita elektrodoým pufrem (viz sekce 3.1.1.). K jednotlivým vzorkům byl přidán vzorkový puf (viz sekce 3.1.1.). Takto připravené vzorky byly zahřáty na 98 °C po dobu 5 minut. Vzorky a marker byly naneseny do jamek v zaostřovacím gelu. Elektroforéza byla připojena ke zdroji napětí 400 V po dobu přibližně 20 minut. Elektroforéza byla ukončena, jakmile modrá barva doputovala ke spodnímu okraji separačního gelu. Gel byl barven negativním barvením popsáním v sekci 3.5.2. nebo vůbec barven nebyl.

3.5.2. Negativní barvení SDS-PAGE gelu

Gel byl nejprve inkubován 10 minut v 0,2M roztoku imidazolu. Poté byl gel převeden do 0,3M roztoku síranu zinečnatého ve kterém byl ponechán, dokud nebyl elektroforetogram zřetelně vidět. Barvení bylo ukončeno promytím v deionizované vodě. Pokud byl gel přenášen na membránu, byl odbarven 2% kyselinou citronovou, třikrát po dobu tří minut⁷⁰.

3.5.3. Elektropřenos proteinů na nitrocelulózovou membránu s chemiluminiscenční detekcí protilátky (Western blot)

3.5.3.1. Elektropřenos

Při elektropřenosu metodou Western blot byly proteiny nejprve separovány metodou SDS elektroforézy popsanou výše (sekce 3.5.1.). Poté byla sestavena aparatura pro elektropřenos a gel byl přichycen k nitrocelulózové membráně (7,5 x 5 cm) tak, aby mezi vrstvami nebyl žádný vzduch, a upevněn v aparatuře. Byla použita aparatura pro polosuché blotování (z angl. Semi-dry blotting). Všechny komponenty byly nejprve namočený v přenosovém pufru, poté byla aparatura uzavřena a připojena ke zdroji napětí 25 V po dobu 30 minut. Pro kvantifikaci produkce jednotlivých knihoven byla použita Bio-Rad aparatura pro mokré blotování (z angl. Wet blotting). Tato aparatura byla naplněna přenosovým puffrem (složení viz sekce 3.1.1.) a připojena ke zdroji napětí 100 V po dobu jedné hodiny.

3.5.3.2. Detekce

Následně byla nespecifická místa na membráně blokována v roztoku BSA (viz sekce 3.1.1.) po dobu jedné hodiny při teplotě 23 °C za stálého míchání. Membrána byla inkubována v roztoku anti-FLAG (10000×ředěna, koncentrace zásobního roztoku 1 mg/ml, Sigma-Aldrich, Německo) či anti-His protilátky (1000×ředěna, koncentrace zásobního roztoku 0,4 mg/ml, Sigma-Aldrich, Německo) ve 3 ml roztoku PBST pufru (viz sekce 3.1.1.). V připraveném roztoku byla membrána inkubována 1 h při teplotě 23 °C za stálého míchání. Po ukončení inkubace byla membrána třikrát po dobu pěti minut promyta v PBST pufru (viz sekce 3.1.1.). Membrána byla vyvolána 600 µl vyvolávacího roztoku SuperSignal West Femto chemiluminiscent solution (ThermoFisher Scientific, USA). Membrána byla inkubována s vyvolávacím roztokem po dobu 5 minut. Nakonec byl blot vizualizován pomocí chemiluminescence v přístroji Amersham Imager 600 (GE Healthcare, UK).

3.5.4. Horizontální agarosová elektroforéza

1% agarosový gel byl připraven rozpuštěním 0,3 g agarosy ve 30 ml 1×koncentrovaného TBE (viz sekce 3.1.1.) pufru. Aby bylo dosaženo rozpuštění agarosy byla směs zahřáta v mikrovlnné troubě. Poté byla směs ochlazená na teplotu asi 50 °C a byly přidány 3 µl fluorescenčního barviva GelRed (BIOTIUM, USA, 10000×ve vodě). Směs byla promíchána a nalita na elektroforetickou desku. Do tekuté směsi nalité na desku byl vložen elektroforetický hřeben pro vytvoření příslušných jamek k nanesení vzorků. Gel tuhnul při pokojové teplotě. Následně byl hřeben vyjmut, gel byl vložen do elektroforetické vany a přelit 1× TBE pufrům po rysku (viz sekce 3.1.1.). Vzorky nanášené na tento gel byly vždy smíchány s nanášecím pufrům 1×GelPilot Loading Dye (Qiagen, Německo). Elektroforéza probíhala při napětí 120 V a byla ukončena, když vzorky putující čelem doputovaly do tří čtvrtin délky gelu. Po ukončení elektroforézy byl gel umístěn na UV transiluminátor a pod UV světlem při 312 nm byly vizualizovány DNA fragmenty.

3.5.5. Akrylamidová denaturující elektroforéza

Pro provedení akrylamidové elektroforézy byla nejprve sestavena aparatura a poté byl připraven gel složený z 6 ml zásobního roztoku (viz sekce 3.1.1.) pro akrylamidovou denaturující elektroforézu (viz sekce 3.1.1., 10 µl TEMED a 60 µl 10% APS). Tento gel byl nalit až po okraj mezi dvě očištěná skla umístěná v držáku a nakonec do něj byl okamžitě ponořen elektroforetický hřeben. Po ztuhnutí gelu byl hřeben vyjmut a skla umístěna do elektroforetické aparatury a převrstvena 1× TBE pufrům (viz sekce 3.1.1.). Do krajní jamky byl nanesen Low Range ssRNA marker (NEB, USA). K jednotlivým vzorkům byl přidán RNA loading Dye (NEB, USA) a také byly naneseny do jamek v gelu. Elektroforéza běžela při napětí 260 V a byla ukončena přibližně po 20 minutách, jakmile modrá barva doputovala ke spodnímu okraji separačního gelu. Poté byl gel přenesen do 50 ml 1×koncentrovaného TBE pufru a bylo k němu přidáno 5 µl fluorescenčního barviva GelRed (BIOTIUM, USA, 10000×ve vodě). Po patnáctiminutové inkubaci byl umístěn na UV transiluminátor a pod UV zářením při 312 nm byly pozorovány RNA fragmenty.

4. Výsledky

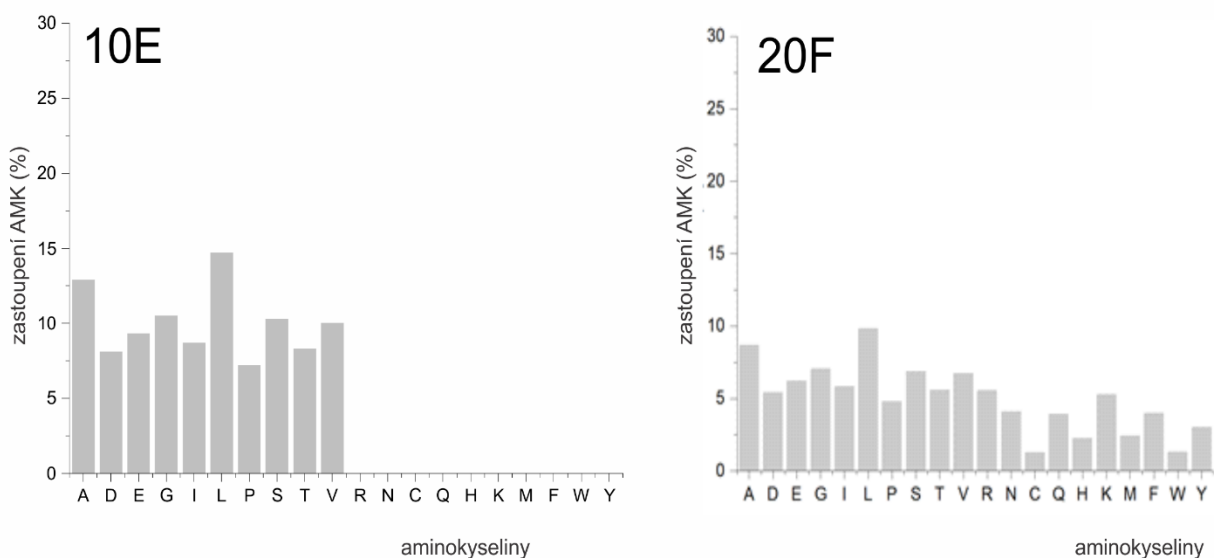
4.1. Příprava DNA a RNA knihoven

Tato práce se zabývá dvěma proteinovými kombinatoriálními knihovnami s různým aminokyselinovým složením. Knihovna označovaná jako 10E je složena pouze z aminokyselin

raných, knihovna označovaná jako 20F je složena ze všech dvaceti proteinogenních aminokyselin (viz. Tab. 15). Rozdělení aminokyselin na rané a pozdní je detailněji popsáno v teoretickém úvodu této práce, viz sekce 1.3. Poměry mezi jednotlivými aminokyselinami v knihovnách, které reflektují jejich přirozené poměrové zastoupení v biologických proteinech, jsou znázorněny na Obr. 6.

Tab. 15 Popis jednotlivých knihoven

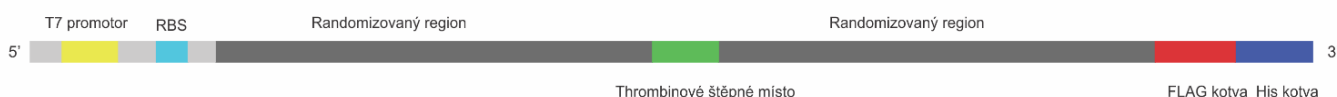
Označení	Kritérium výběru aminokyselin	Aminokyseliny
20F	Všech dvacet proteinogenních aminokyselin (z angl. Full)	G, A, D, V, P, S, E, T, L, R, N, I, Q, H, K, C, F, Y, M, W
10E	Deset raných aminokyselin (z angl. Early)	A, D, E, G, I, L, P, S, T, V



Obr. 6 Procentuální zastoupení jednotlivých aminokyselin v knihovnách. Převzato a upraveno⁶⁵.

Obě knihovny byly nejprve navrženy ve formě randomizovaných jednořetězcových DNA oligonukleotidů, které končí nebo začínají definovanou sekvencí. Uvedená definovaná sekvence slouží jednak ke spojení dvou oligonukleotidů v tomto místě, protože bylo dosaženo limitní délky komerční syntézy a následně také k charakterizaci tendence tvorby terciární struktury sekvencí v knihovnách (viz sekce 4.2.6.). Randomizované regiony byly navrženy pomocí algoritmu CoLiDe⁶⁵ vyvinutého v laboratoři, kde byla tato práce vypracována a výsledná variabilita takto navržených knihoven odpovídá přibližně 10^{11} . Tyto oligonukleotidy byly syntetizovány komerčně firmou Integrated DNA Technologies (IDT, USA). Schématické znázornění jednotlivých sekvencí

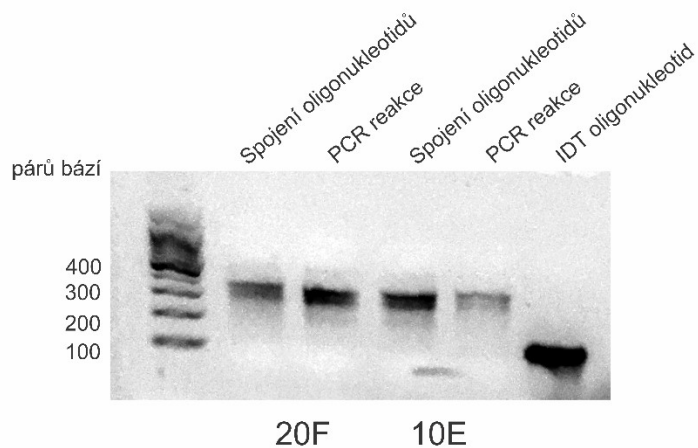
v knihovnách po spojení obou oligonukletidů, dosyntetizování zbytku řetězce a přidání His-kotvy na 3' konec (více viz sekce 3.2.1. a 3.2.2.) je znázorněno na Obr. 7. Všechny popisované sekvence mají na svém začátku T7 promotor pro nasednutí T7 polymerázy, která dané sekvence transkribuje do mRNA molekul. Následuje místo pro nasednutí ribozomu a poté náhodný úsek o délce 277 bází, který je v polovině přerušen definovanou sekvencí, jejíž účel byl již popsán výše v tomto odstavci. Na svém konci sekvence obsahují FLAG kotvu, která slouží k purifikaci a následné detekci knihoven pomocí metody Western blot. Délka jednotlivých sekvencí v knihovnách po spojení obou oligonukleotidů, před přidáním His-kotvy, je 374 bází.



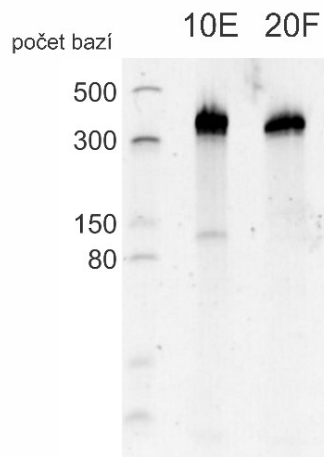
Obr. 7 Schematické znázornění sekvencí v knihovnách po přidání histidinové kotvy pomocí PCR reakce

Oba komerčně syntetizované oligonukleotidy od jednotlivých knihoven byly spojeny v definovaném místě překryvu a následně byla chybějící část řetězce dosyntetizována pomocí Klenow polymerázy od místa překryvu do konce řetězce (viz sekce 3.2.2.). V dalším kroku byla pomocí PCR reakce do jednotlivých sekvencí přidána His-kotva na 3' konec (viz sekce 3.2.3.). DNA produkty těchto kroků, znázorněny na Obr. 8, se zdají být mírně rozmyté, což odpovídá charakteru knihovny s velkým množstvím lišících se sekvencí. Pokud porovnáme produkty uvedené reakce mezi oběma knihovnami (viz Obr. 8), nejsou patrné žádné zásadní rozdíly mezi produkty jednotlivých knihoven. Následně byla provedena PCR reakce z důvodu přidání histidinové kotvy do sekvencí v obou knihovnách. V tomto případě se zdá, že výtěžek této reakce byl nepatrně větší u knihovny 20F (viz Obr. 8), což se potvrdilo i po spektrofotometrickém změření koncentrace purifikovaného produktu reakce.

Obě knihovny se poté podařilo transkribovat. Transkripce knihoven do formy mRNA byla provedena pomocí komerčního kitu Ampliscribe T7 Flash Transcription kitu (Lucigen, USA). Na Obr. 9 lze znovu pozorovat mírně rozmyté produkty reakce, což jak již bylo řečeno výše odpovídá charakteru knihovny, takže lze předpokládat alespoň částečné zachování variability v sekvencích. Pokud porovnáme obě knihovny mezi sebou, lze vidět, že reakce proběhly téměř bez rozdílů ve výsledném výtěžku (viz Obr. 9), což se znovu potvrdilo i spektrofotometricky. Výsledná mRNA byla následně použita jako templát pro translační reakci.



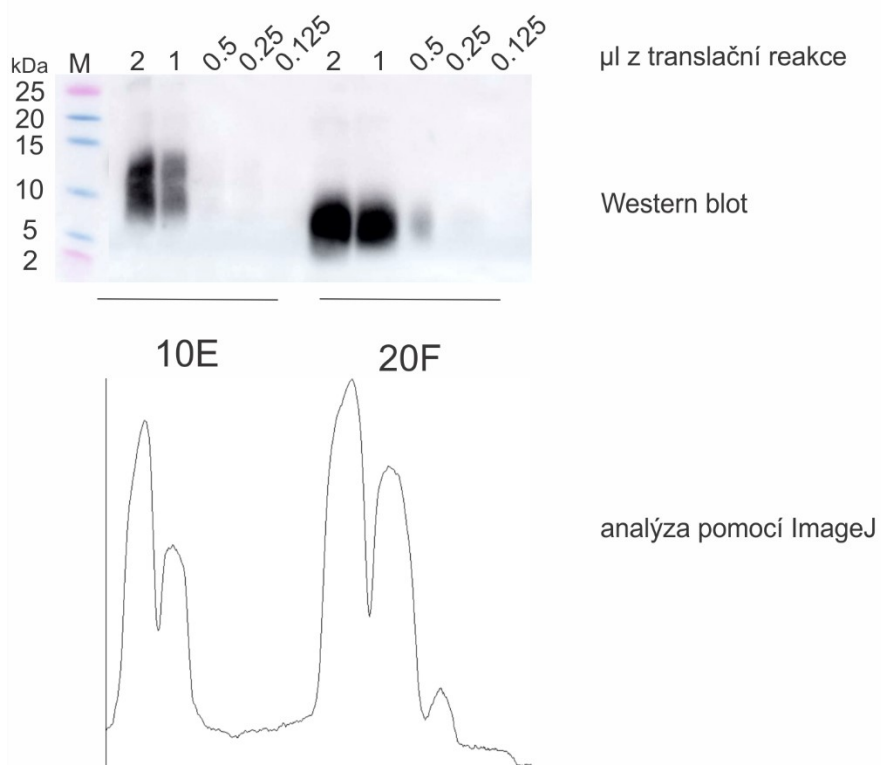
Obr. 8 Analýza DNA knihoven. 1% agarosový gel znázorňující knihovny 10E a 20F po spojení oligonukleotidů a přidání histidinové kotvy. Zleva: 3 μ l markeru 100 bp DNA ladder (NEB, USA), 20F knihovna po spojení oligonukleotidů, 20F knihovna po PCR reakci, 10E knihovna po spojení oligonukleotidů, 10E knihovna po PCR reakci, IDT oligonukleotid



Obr. 9 Analýza RNA knihoven. Akrylamidový denaturující gel s RNA od obou knihoven. Zleva: 2,5 μ l low range ssRNA markeru; 500 ng RNA z příslušné knihovny

4.2. Translace knihoven

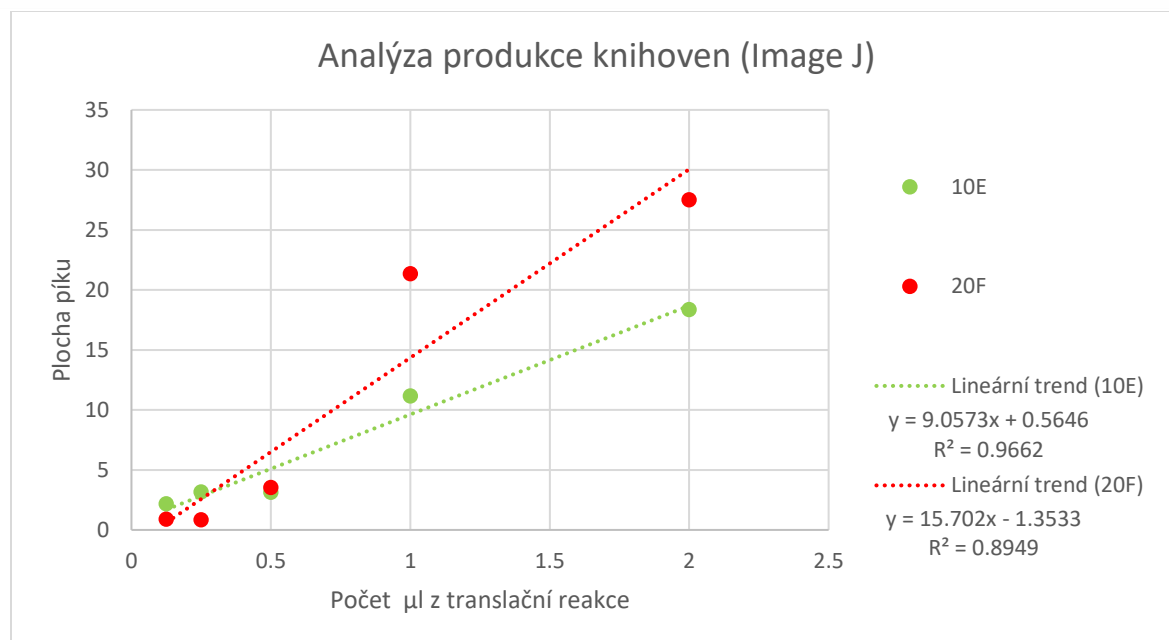
Translace knihoven byla provedena pomocí bezbuněčné exprese s využitím komerčního kitu PUREfrex 2.0 (GeneFrontier, Japonsko). Jako reakční templát byla použita mRNA daných knihoven, jejíž syntéza je popsána v předchozím kroku (viz sekce 4.1.1.). Detailněji je uvedený postup uveden v Experimentálních metodách v sekci 3.2.2. Míra produkce obou proteinových knihoven je znázorněna na Obr. 10. Pro účely analýzy produkce knihoven byla vždy pipetována $\frac{1}{4}$ standardní Purefrex reakce (5 μ l) a poté příslušně rozdělena, jak je patrné z Obr. 10. Analýza produkce proteinů byla provedena pomocí metody Western blot (viz sekce 3.5.3.). Následně byl obrázek získaný metodou Western blot zpracován pomocí programu Image J, ve kterém byly analyzovány intenzity skvrn, které odpovídají knihovnám. Z ukázaných experimentálních výsledků vztahujících se k účinnosti translace a jejich analýzy (viz Obr. 10, Tab. 16, Obr. 11) je patrné, že knihovna 10E je produkována s přibližně o jednu třetinu nižší účinností než knihovna 20F (spočítáno pro dvě nejvyšší nanášky, menší nanášky nevykazují markantnější rozdíly). Data sesbíraná pro knihovnu 10E se také více blíží předpokladu o lineárnosti analyzovaných intenzit (viz Obr. 11). Širší spektrum produktů s různou molekulovou hmotností patrné z Obr. 10 pravděpodobně odráží různé aminokyselinové složení peptidů v knihovně. Toto pozorování lze považovat za důkaz, že alespoň částečná navržená variabilita knihoven zůstala zachována i po translačním procesu.



Obr. 10 Analýza účinnosti translace. Western blot (mokré blotování, Novex (10-20%) tricinový gel (ThermoFisher Scientific, USA) obsahující různá množství translační reakce od obou knihoven, analýza produkce pomocí programu Image J. Zleva: Precision Plus Protein Dual Color Standard - 3 µl, pět různých množství translační reakce od knihovny 10E, pět různých množství translační reakce od knihovny 20F. Vizualizace chemiluminescence, anti-FLAG protilátka.

Tab. 16 Plochy píků ve spektru znázorněném na Obr. 10 a jejich kvantitativní srovnání

počet μ l z translační reakce	plocha píku (Image J)	podíl vzhledem k největšímu píku	knihovna
2	18.34	0.67	10E
1	11.15	0.41	
0.5	3.12	0.11	
0.25	3.14	0.11	
0.125	2.17	0.08	
2	27.5	1.00	20F
1	21.34	0.78	
0.5	3.52	0.13	
0.25	0.84	0.03	
0.125	0.88	0.03	

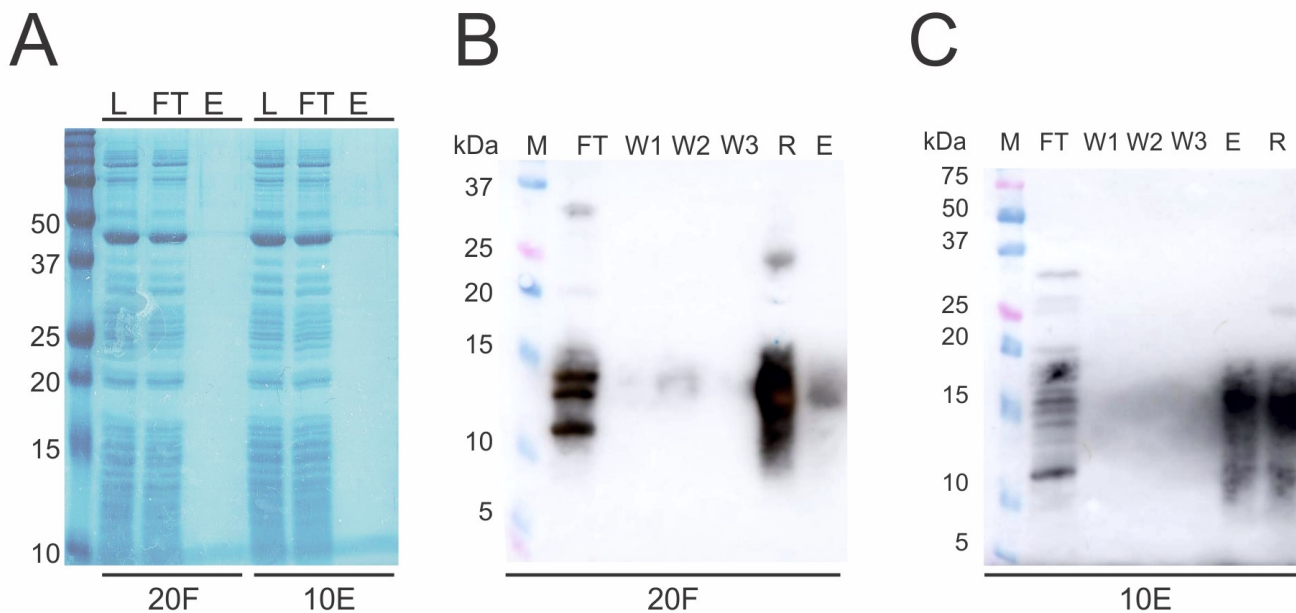


Obr. 11 Analýza intenzity produkce z western blotu pomocí programu Image J. Na grafu jsou vyneseny plochy píků znázorněných na Obr. 8 v závislosti na počtu μ l z translační reakce. Tečkované přímky zobrazené v grafu ukazují trend intenzit píků stoupat lineárně (vzhledem ke zvolené lineárně rostoucí kvantitě zobrazené translační reakce).

4.3. Purifikace proteinových knihoven

Pro účely strukturní analýzy byly proteinové knihovny 10E a 20F purifikovány afinitní chromatografií s využitím FLAG kotvy (viz sekce 3.2.7.), která se nachází na 3' konci sekvencí před kotvou histidinovou. Bezbuněčnou expresí komerčním kitem bylo možné získat jen velmi malé množství knihoven (odhadem 10-100 ng z 20 μ l reakce) a spektrum možných biofyzikálních

metod pro jejich analýzu tak bylo velmi omezené. Průběh purifikace byl sledován pomocí SDS-PAGE elektroforézy, přičemž detekce proteinů probíhala méně tradiční metodou, negativním barvením⁷⁰ (viz sekce 3.5.2.). Tento způsob barvení má sice nižší detekční limit v porovnání s běžnými metodami barvení, nicméně není založen na specifickém aminokyselinovém složení a mohla by tak být detekována i knihovna 10E, kde nejsou zastoupeny např. kladně nabitě a aromatické aminokyseliny. Ve většině případů nebyly knihovny po obarvení gelu touto metodou viditelné (viz Obr. 12 část A). Obě knihovny se podařilo purifikovat. Celý purifikační proces procházel dlouhou optimalizací, a i přesto při něm stále docházelo k velkým ztrátám. V procesu purifikace jsou patrné výrazné rozdíly mezi knihovnami. Z Obr. 12 (část B, část C) je viditelné, že se lépe purifikovala knihovna 10E. Většina proteinů z knihovny 20F zůstává přichycena na afinitním FLAG nosiči i po provedení elučního kroku, u knihovny 10E je tomu tak také, ale pouze asi u poloviny produktů translační reakce. Pro některé analýzy, které nevyžadovaly nativní vzorek, byla využívána také purifikace denaturační, aby bylo získáno větší množství produktu pro analýzu. Detekčního množství pro plánovanou aminokyselinovou analýzu za účelem kontroly zastoupení jednotlivých aminokyselin v knihovnách se nepodařilo u těchto vzorků dosáhnout.



Obr. 12 Purifikace proteinových knihoven

A 18% akrylamidový gel z SDS-PAGE elektroforézy obsahující frakce z afinitní purifikace obou knihoven metodou s využitím FLAG nosiče. Barven negativním barvením. Zleva: Precision Plus Protein Dual Color Standard - 3 µl, 1 µl translační reakce (L), 1 µl frakce z translační reakce nenavázané na nosič (FT), 10 µl eluční frakce (E) od knihovny 20F, následně stejně u knihovny 10E

B Western blot (polosuché blotování, 18% akrylamidový gel z SDS-PAGE elektroforézy) obsahující vzorky z purifikace pomocí FLAG nosiče od knihovny 20F. Zleva: 3 µl Precision Plus Protein Dual Color Standard, FT – 5 µl směsi translační reakce nenachytané na nosič, W1, W2, W3 – 10 µl jednotlivých promývacích frakcí, R – FLAG nosič po povaření v 10 µl 1×vzorkovém pufru, E – 10 µl eluční frakce. Vizualizace chemiluminescence, anti-FLAG protilátka.

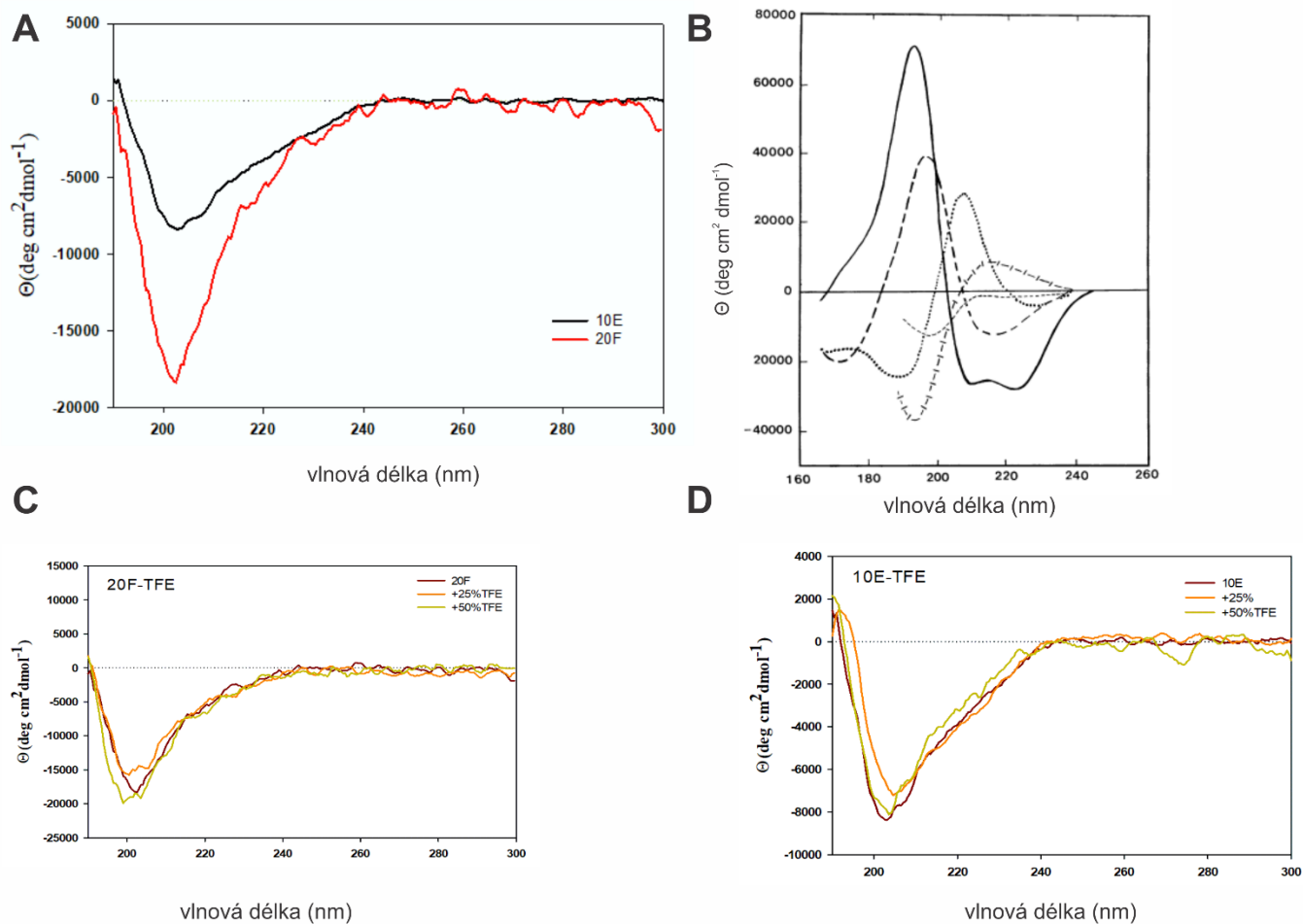
C Western blot (polosuché blotování, 18% akrylamidový gel z SDS-PAGE elektroforézy) obsahující vzorky z purifikace pomocí FLAG nosiče od knihovny 10E. Zleva: Precision Plus Protein Dual Color Standard - 3 µl, FT – 5 µl směsi translační reakce nenachytané na nosič, W1, W2, W3 – 10 µl jednotlivých promývacích frakcí, E – 10 µl eluční frakce, R – FLAG nosič po povaření v 10 µl 1×vzorkovém pufru. Vizualizace chemiluminescence, anti-FLAG protilátka.

4.5. CD spektroskopie proteinových knihoven

Vzorky pro účely měření CD spekter byly z DNA knihoven připraveny pomocí transkripce a translace (v objemu 500 μ l translační reakce) a následně purifikovány pomocí FLAG afinitního nosiče pouze s rozdílem v použitém eluční pufru (viz sekce 3.3.2., Obr. 13.). V knihovnách nebyl potvrzen žádný dominantní motiv sekundární struktury. Numerická analýza dat (viz Tab. 17) ukázala pro knihovnu 10E minoritní zastoupení helikální struktury a porovnatelného podílu struktury skládaného listu, otoček a neuspořádané struktury (α -helix 9 %, β -skládaný list 33 %, β -ohyb 25 %, nestrukturovaný 33 %) a pro 20F porovnatelný podíl v zastoupení helikální struktury, struktury skládaného listu a otoček a největší podíl neuspořádané struktury (α -helix 21 %, β -skládaný list 22 %, β -ohyb 24 %, nestrukturovaný 33 %) (viz část A a část B na Obr. 13). Titrace pomocí TFE dále ukázala, že knihovny nemají preferenci pro zvýšenou tvorbu helikální struktury (část C a část D na Obr. 13). Koncentrace vzorku pro snímání těchto spekter byla hraniční.

Tab.17 Numerická analýza dat z CD spekter (z Obr. 13, část A)

Parametr	10E	20F
α -helix	0.090	0.210
β -skládaný list	0.330	0.220
β -ohyb	0.250	0.240
nestrukturovaný	0.330	0.330
odchylka (NRMSD)	0.120	0.198
koncentrace [M/l]	0.0111	0.0036



Obr. 13 CD spektra proteinových knihoven

A CD spektra knihoven 10E (černá) a 20F (červená)

B Modelová spektra různých strukturních modelů proteinů. α -helix (nepřerušovaná čára), anti-paralelní β -skládaný list (dlouze přerušovaná čára), β -ohyb (tečkovaná čára), 3-helix nebo poly(Pro) (křížem přerušovaná čára), nestrukturovaný protein (krátce přerušovaná čára). Převzato a upraveno⁷¹.

C Titrace pomocí TFE (knihovna 20F)

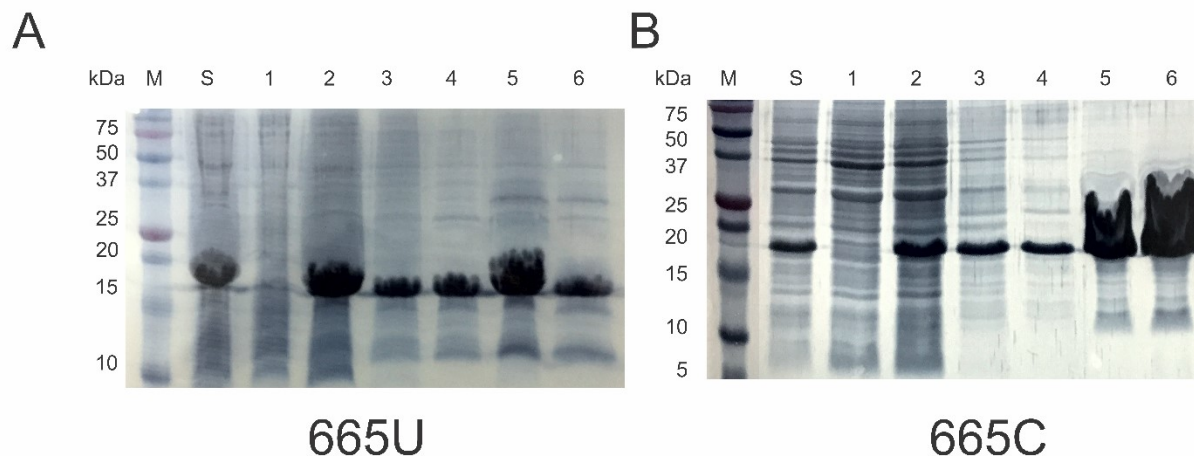
D Titrace pomocí TFE (knihovna 10E)

4.6. Charakterizace tvorby terciární struktury

Charakterizace potenciálu pro tvorbu terciární struktury u obou knihoven byla analyzována pomocí proteolytického štěpení enzymem thrombinem. Jak již bylo uvedeno v popisu návrhu jednotlivých sekvencí v knihovnách (viz sekce 4.1.1.), všechny jednotlivé sekvence obsahují ve svém středu definovanou nenáhodnou sekvenci (aminokyselinová sekvence LVPRGS), kterou tento enzym rozpoznává a štěpí. Předpokládáme, že pokud proteiny v dané knihovně mají tendenci tvořit sbalenou terciární strukturu, tato definovaná sekvence bude s větší pravděpodobností ukryta uvnitř sbalených proteinů a tím pádem pravděpodobnost, že proteiny v knihovně budou enzymem štěpeny bude nižší než u proteinů nesbalených, netvořících terciární strukturu⁵⁴. Pro ověření specifity trombinového štěpení byla metodika nejdříve ověřena na syntetickém neuspořádaném proteinu, jak je popsáno v následující sekci.

4.6.1. Kontrolní proteiny pro průkaznost specifity štěpení enzymu thrombinu

Kontrolní proteiny použité v této práci byly vybrány na základě předchozího výzkumu skupiny jejíž součástí je i tato práce^{59,69}. Protein s označením 665U patří do skupiny proteinů známých jako NBPs (viz sekce 1.4.3.) a byl charakterizován jako dobře se produkující v bakterii *E. coli*, solubilní a nestrukturovaný. Tento protein neobsahuje ve své sekvenci thrombinové štěpné místo. Protein označovaný jako 665C je z velké části kopií sekvence proteinu 665U, část této sekvence však byla nahrazena specifickou sekvencí, kterou rozpoznává a štěpí enzym thrombin používaný k charakterizaci tvorby terciární struktury proteinových knihoven (viz sekce 3.3.3.). Sekvence proteinu 665U již byla připravena a vnesena do vektoru pET24a dříve, díky předchozímu výzkumu skupiny. Sekvence proteinu 665C byla do tohoto vektoru vložena během experimentální části této práce (viz sekce 3.4.1. až 3.4.4.). Poté byly tyto proteiny exprimovány v bakterii *E. coli* kmen BL21 (viz sekce 3.4.5.), a následně sonikovány (viz sekce 3.4.7.) a purifikovány pomocí afinitní purifikace s využitím histidinové kotvy (viz sekce 3.4.8.). Výsledek afinitní purifikace obou proteinů je znázorněn na Obr. 14, kde na SDS-PAGE gelu pozorujeme větší pruh o velikosti o něco větší, než je vypočítaná molekulová hmotnost proteinů, což odpovídá nestrukturovanému proteinu. Vypočítaná molekulární hmotnost proteinů odpovídá hodnotám 12,13 kDa (665U) a 12,04 kDa (665C). Dále je patrné, že se protein 665C purifikoval lépe. Gely byly barveny znovu metodou negativního barvení (viz sekce 3.5.2.). Koncentrace proteinů po spojení elučních frakcí a zakoncentrování byla změřena spektrofotometricky.



Obr. 14 Purifikace kontrolních proteinů

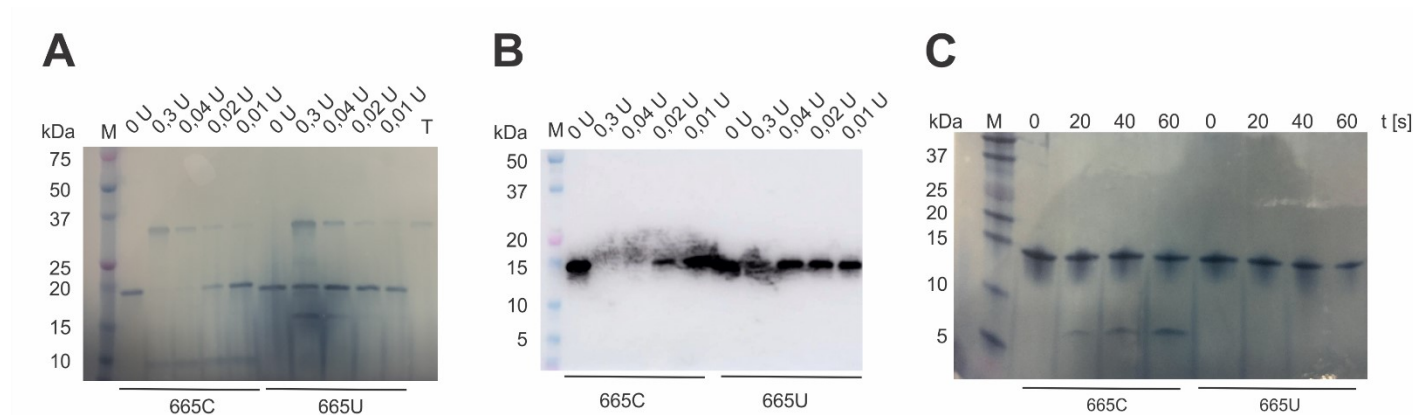
A 18% akrylamidový gel z SDS-PAGE elektroforézy obsahující frakce z afinitní purifikace proteinu 665U. Zleva: 3 μ l Precision Plus Protein Dual Color Standard (M), 5 μ l sonikátu po centrifugaci (S), 5 μ l sonikátu po prolití purifikační kolonou (1), 10 μ l promývací frakce (2–4), 10 μ l eluční frakce (5,6). Očekávaná velikost proteinu 665U: 12,13 kDa.

B 18% akrylamidový gel z SDS-PAGE elektroforézy obsahující frakce z afinitní purifikace proteinu 665U. Zleva: 3 μ l Precision Plus Protein Dual Color Standard (M), 5 μ l sonikátu po centrifugaci (S), 5 μ l sonikátu po prolití purifikační kolonou (1), 10 μ l promývací frakce (2–4), 10 μ l eluční frakce (5,6). Očekávaná velikost proteinu 665C: 12,04 kDa.

4.6.2. Optimalizace podmínek štěpení

S kontrolními proteiny 665C a 665U byla provedena optimalizace štěpení enzymem thrombinem kvůli prokázání specifity štěpení tohoto enzymu. Optimalizace byla provedena za použití různých koncentrací daného enzymu a také lišících se dob štěpení. Štěpení vždy probíhalo při teplotě 37 °C. Během optimalizace koncentrace použitého enzymu reakce běžely po dobu 30 minut. Při optimalizaci doby štěpení bylo použito 0,01 U enzymu na 10 μ g daného proteinu. Očekávaná velikost fragmentů proteinu 665C po úspěšném specifickém štěpení je 6,4 kDa a 5,65 kDa. Výsledky popsanych optimalizací jsou znázorněny na Obr. 15. Vnitřně neuspořádaný kontrolní protein 665C, který ve svém středu obsahuje thrombinové štěpné místo, se ukázal jako dobře štěpitelný v poměrně krátkém čase a při nízkých koncentracích enzymu. Na Obr. 15 je v obou případech optimalizací s rostoucí koncentrací enzymu, či se zvyšující se dobou štěpení na SDS-PAGE gelu (část A a část C), vidět úbytek tohoto proteinu a objevující se produkt štěpení

s očekávanou velikostí. Vnitřně neuspořádaný kontrolní protein 665U, který ve svém středu neobsahuje thrombinové štěpné místo, se při nízkých koncentracích enzymu a při krátkém trvání reakce ukázal jako neštěpený, avšak při koncentraci enzymu 0,3 U na 10 µg proteinu (viz Obr.15, část A) se objevuje fragment, naznačující nespecifické štěpení tohoto proteinu. Proteolytické fragmenty se nepodařilo identifikovat pomocí metody Western blot, jak je patrné z Obr. 15, část B.



Obr. 15 Optimalizace podmínek štěpení na kontrolních proteinech (Očekávaná velikost proteinu 665U: 12,13 kDa. Očekávaná velikost proteinu 665C: 12,04 kDa. Očekávaná velikost fragmentů proteinu 665C: 6,4 kDa a 5,65 kDa.)

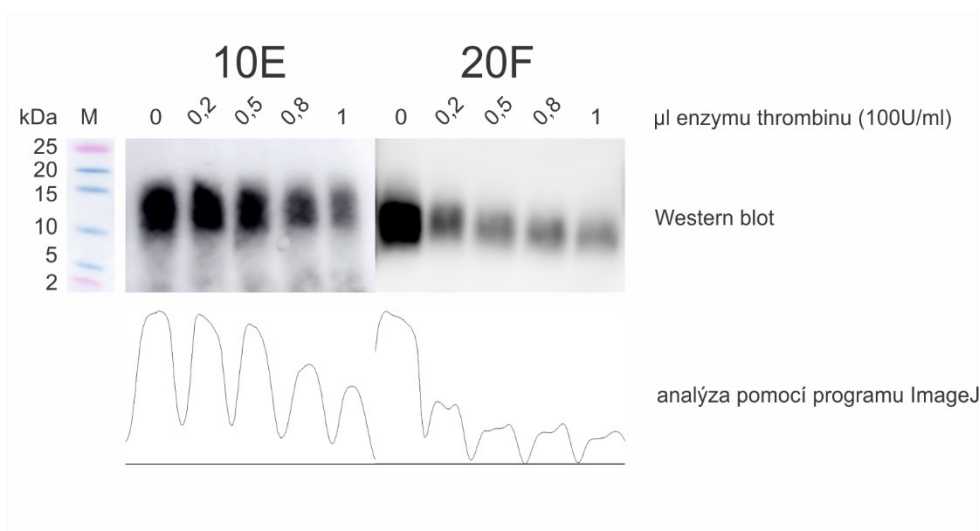
A Gradientový gel (8-16%) Mini-PROTEAN TGX (Bio-Rad, USA) z SDS-PAGE elektroforézy obsahující optimalizaci štěpení proteinů 665C a 665U. Zleva: 3 µl Precision Plus Protein Dual Color Standard (M), 10 µl reakce se čtyřmi různými koncentracemi enzymu thrombinu (koncentrace vztažené na 10 µg daného proteinu) pro 665C a poté pro 665U, samotný enzym (T).

B Western blot obsahující optimalizaci štěpení proteinů 665C a 665U. Zleva: 3 µl Precision Plus Protein Dual Color Standard (M), 10 µl reakce se čtyřmi různými koncentracemi enzymu thrombinu (koncentrace vztažené na 10 µg daného proteinu) pro 665C a poté pro 665U. Vizualizace chemiluminiscencí, anti-His protilátka.

C Tricinový gel (16,5%) Mini-PROTEAN Tris-Tricine (BioRad) obsahující časovou optimalizaci štěpení proteinů 665C a 665U. Zleva: 3 µl Precision Plus Protein Dual Color Standard (M), 10 µl reakce ukončené po čtyřech různých časech pro 665C a poté pro 665U. Koncentrace enzymu 0,01 U na 10 µg daného proteinu.

4.6.3. Štěpení kombinatoriálních knihoven enzymem thrombinem

Po testování specifity a účinnosti štěpení na kontrolních proteinech byla tato metoda aplikována i na kombinatoriální knihovny. Postup štěpení je detailněji popsán v sekci 3.3.3. Znovu byly provedeny optimalizace podmínek z hlediska množství použitého enzymu, štěpení probíhalo vždy při 37 °C po dobu 30 minut. Na jednu proteolytickou reakci (10 µl eluční frakce) byla vždy použita 1/5 translační reakce (20 µl) po afinitní purifikaci na FLAG nosiči (celá eluční frakce 50 µl, viz sekce 3.2.7.) a bylo přidáno uvedené množství enzymu. Koncentrace použitého enzymu je v tomto případě vzhledem k odhadnutému množství knihoven (2-20 ng na jednu proteolytickou reakci) mnohem vyšší než u výše uvedeném štěpení na kontrolních proteinech. Analýza štěpení byla provedena pomocí metody Western blot (viz sekce 3.5.2.) a následném zpracování pomocí programu Image J. Z Obr. 16 lze odvodit, že obě knihovny podléhají proteolytickému štěpení. Z dat uvedených v Tab. 18 vyplývá, že knihovna 20F podléhá štěpení použitým enzymem ve větší míře než knihovna 10E. U této knihovny nepozorujeme výraznější rozdíly při vyšších množstvích použitého enzymu. Množství nenaštěpených sekvencí 20F odpovídá přibližně 20 % (vypočítané vzhledem k největšímu píku negativní kontroly) téměř u všech optimalizací množství enzymu na Obr.16. Míra naštěpených sekvencí u knihovny 10E narůstá se zvyšující se koncentrací enzymu, při nejvyšším použitém množství enzymu zůstává nenaštěpených přibližně 36 % sekvencí (znovu vypočítané vzhledem k největšímu píku negativní kontroly). Produkt proteolytického štěpení knihoven se na Western blotu nepodařilo identifikovat. Očekávaná velikost fragmentů naštěpených knihoven je okolo 5 kDa.



Obr. 16 Analýza míry štěpení. Western blot (mokrý blotování, Novex (10-20%) tricínový gel (ThermoFisher Scientific, USA) obsahující vždy 10 µl eluční frakce (1/5 translační reakce) dané knihovny a různá množství enzymu thrombinu, analýza štěpení pomocí programu Image J. Zleva: Pět štepných reakcí (vždy 1/5 translační reakce) s knihovnou 10E, lišících se množstvím použitého enzymu, pět štepných reakcí (vždy 1/5 translační reakce) s knihovnou 20F, lišících se množstvím použitého enzymu. Vizualizace chemiluminescence, anti-FLAG protilátka.

Tab. 18 Plochy píků ve spektru znázorněném na Obr.17 a jejich kvantitativní srovnání

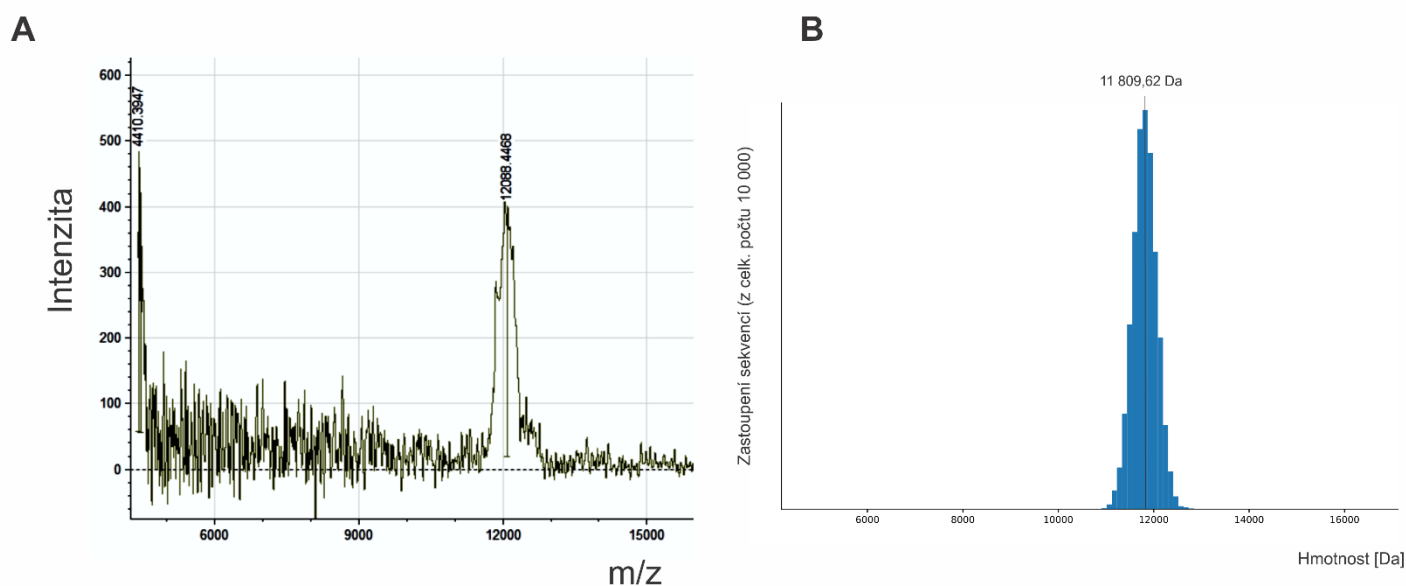
počet µl použitého enzymu	plocha píku (Image J)	podíl vzhledem k největšímu píku (negativní kontrola)	knihovna
0	28.27	1.00	10E
0.2	24.11	0.85	
0.5	21.76	0.77	
0.8	15.82	0.56	
1	10.04	0.36	
0	52.35	1.00	20F
0.2	16.66	0.32	
0.5	11.30	0.22	
0.8	10.31	0.20	
1	9.38	0.18	

4.7. Hmotnostní analýza proteinových knihoven (technika MALDI-TOF)

Pro ověření správnosti vzorků a heterogenity obou knihoven bylo připraveno větší množství 10E a 20F knihoven denaturační purifikací s využitím histidinové kotvy (viz sekce 3.2.8.). Pro účely této analýzy nebylo třeba nativního vzorku a tento způsob purifikace poskytoval výrazně větší výtěžky (odhadem 1 µg z 20 µl translační reakce). Hmotnostní spektra byla změřena na obou knihovnách pomocí techniky MALDI-TOF. Tato spektra uvedená na Obr. 17 a Obr. 18 v části A jsou důkazem o přítomnosti daných knihoven v analyzovaném roztoku. V části B uvedených

obrázků jsou znázorněny distribuce předpokládaných hmotností pro deset tisíc proteinových sekvencí v daných knihovnách. Pokud porovnáme průměrnou velikost knihoven z naměřených spekter s předpokládanou průměrnou velikostí, vidíme, že naměřená hodnota je u obou knihoven nepatrně vyšší.

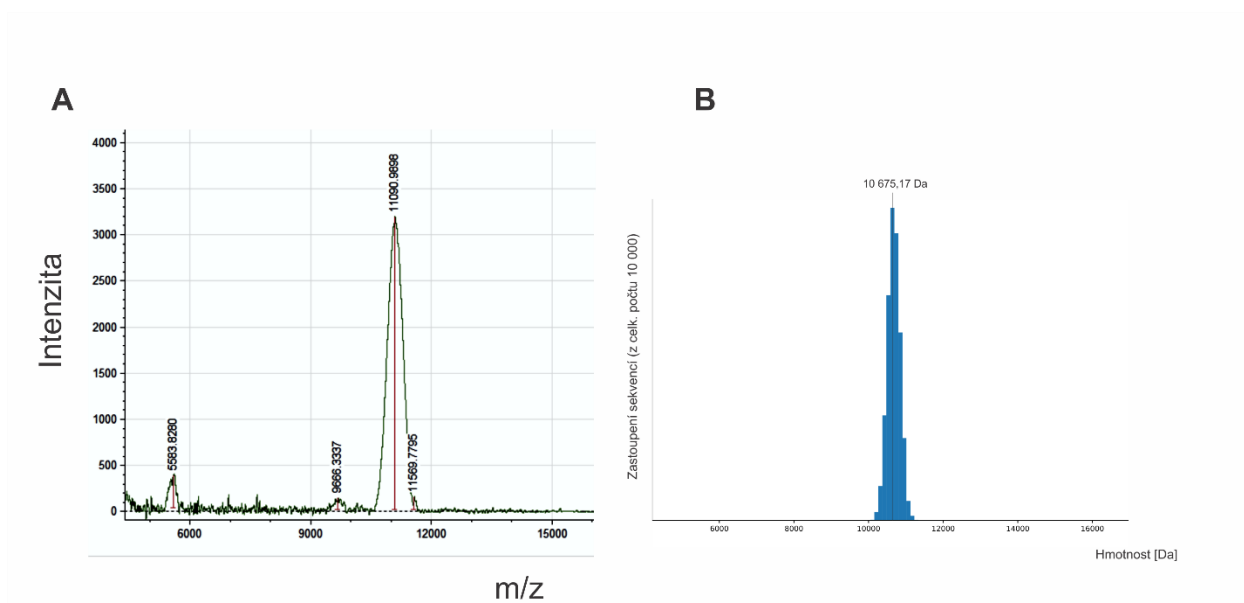
Široký pík poukazuje na přítomnost většího množství lišících se sekvencí, a ne pouze sekvence jediné, protože různé sekvence se mírně liší ve své molekulové hmotnosti vzhledem k lišícím se aminokyselinám. Knihovna 20F má dle spektra molekulovou hmotnost o něco větší, což odpovídá většímu zastoupení aminokyselin s větší molekulovou hmotností v této knihovně a také předpokládaným distribucím hmotností náhodných sekvencí.



Obr. 17 Srovnání skutečné a predikované hmotnosti knihovny 20F

A Hmotnostní spektrum naměřené na knihovně 20F. Graf zobrazuje intenzitu (svislá osa) v závislosti na poměru hmotnosti a náboje (vodorovná osa). Proteinové knihovně odpovídá pík o velikosti $\approx 11\,091$.

B Distribuce předpokládaných hmotností pro 10 000 náhodných sekvencí z knihovny 20F. Graf zobrazuje předpokládanou hmotnost (vodorovná osa) v závislosti na počtu sekvencí (svislá osa).



Obr. 18 Srovnání skutečné a predikované hmotnosti knihovny 10E

A Hmotnostní spektrum naměřené na knihovně 10E. Graf zobrazuje intenzitu (svislá osa) v závislosti na poměru hmotnosti a náboje (vodorovná osa). Proteinové knihovně odpovídá větší pík o velikosti $\approx 12\,088$.

B Distribuce předpokládaných hmotností pro 10 000 náhodných sekvencí z knihovny 10E. Graf zobrazuje předpokládanou hmotnost (vodorovná osa) v závislosti na počtu sekvencí (svislá osa).

5. Diskuze

Cílem práce byla příprava dvou proteinových kombinatoriálních knihoven s navrženou variabilitou $\sim 10^{11}$ náhodných proteinových sekvencí o délce sto aminokyselin s jediným definovaným úsekem uprostřed náhodného regionu (o délce šest aminokyselin, sloužícím k následné charakterizaci tvorby 3D struktur) a následná charakterizace tendence tvorby sekundárních a terciárních struktur u těchto knihoven. V dřívějších studiích, zabývajících se tematikou tvorby struktury u NBPs, byly vždy vybrány pouze jednotlivé sekvence navržených knihoven, které byly následně charakterizovány z hlediska struktury, solubility či funkce^{52–54,60}. Tato práce se jako první (podle dostupných informací) zabývá charakterizací tendence tvorby sekundárních a terciárních struktur na kombinatoriálních knihovnách jako celku. Většina zmíněných studií pak také k produkci náhodných sekvencí používá bakterie *E. Coli*^{52–54}, tato studie naopak využívá metody bezbuněčné exprese. Použitá metoda této exprese je nevýhodná z hlediska nízkých výtěžků translačních reakcí a pro některé metody nebylo v rámci této práce dosaženo detekčního limitu (např. aminokyselinová analýza). Metoda bezbuněčné exprese však umožňuje produkci a následnou charakterizaci velkého souboru náhodných sekvencí v knihovně jako celku. Navržená knihovna 20F obsahuje dvacet proteinogenních aminokyselin, navržená knihovna 10E obsahuje pouze deset aminokyselin raných (dle studie P. G. Higgse a R. E. Pudritz³⁹). Dříve již byly provedeny studie, zabývající se charakterizací proteinů s redukovanou aminokyselinou abecedou, které využívají mutací existujících proteinů^{48,49}. Tato studie se však zabývá charakterizací náhodných sekvencí s redukovanou aminokyselinou abecedou.

Obě navržené knihovny se podařilo připravit v DNA formě, následně převést do mRNA formy a poté produkovat ve formě proteinů. Výsledná navržená variabilita byla pravděpodobně během tohoto procesu částečně snížena, stále se však jedná o vzorky s velmi vysokou výslednou variabilitou v porovnání s dřívějšími, výše uvedenými studiemi. Při expresi jednotlivých knihoven byly pozorovány vyšší výtěžky produkce knihovny 20F v porovnání s 10E, což potvrzuje trend pozorovaný při produkci jednotlivých sekvencí s podobným aminokyselinovým složením v práci Newton et. al.⁵².

Purifikace obou knihoven procházela náročným procesem optimalizace. Komplikace byly způsobeny vysokou variabilitou vzorků, nízkou účinností samotné translační reakce a pravděpodobně vysokou tendencí sekvencí v obou knihovnách k tvorbě agregátů s proteiny

z expresního systému. Purifikace knihovny 10E probíhala lépe, což naznačuje nižší tendence k agregaci sekvencí této knihovny a zároveň také vyšší solubilitu těchto sekvencí a to je v souladu s dřívějšími studiemi^{52,53}.

Kvalitu a variabilitu produkovaných sekvencí se podařilo ověřit díky hmotnostní spektroskopii, kde pozorujeme široký pík s hmotnostním rozmezím přibližně odpovídajícím teoretickému předpokladu u obou knihoven. V rámci této diplomové práce se nepodařilo dosáhnout množství vzorku pro aminokyselinovou analýzu knihoven pro ověření správného složení knihoven. Zvolený postup přípravy knihoven byl nicméně hlouběji studován kolegy v laboratoři na příkladu jiné proteinové knihovny, u které bylo výsledně správné složení prokázáno na úrovni DNA a mRNA (pomocí “next-generation sequencing”) a rovněž i aminokyselinovou analýzou proteinového produktu⁶⁵. Lze tedy předpokládat, že díky zachovalému postupu při přípravě knihoven bude stejná kvalita kontroly úspěšně provedena i v případě knihoven 20F a 10E v rámci pokračování tohoto projektu.

CD spektra naměřená na obou knihovnách nevykazují žádný dominantní motiv sekundární struktury, což je znovu v souladu se studií Newton et. al.⁵². Můžeme očekávat, že ve velkém souboru náhodných sekvencí bude velká míra neuspořádanosti a rovněž kombinace obou hlavních strukturních motivů. Na základě numerické analýzy dat byl ovšem pozorován zvýšený podíl β -skládaného listu u knihovny 10E (33 % oproti 22 % u knihovny 20F), zatímco u knihovny 20F byl v porovnání s 10E větší podíl α -helikální struktury (21 % vs 9 %). Mírná preference raných proteinů pro β -skládaný list byla predikována na základě dřívějších teoretických studií a byla rovněž pozorována v ribozomálním strukturním jádře^{10,12,13,46,47}. V rámci širšího projektu v laboratoři byla rovněž provedena bioinformatická predikce sekundární struktury na souboru sekvencí vybraných z knihovny 20F a 10E a i tato predikce vyhodnotila větší preferenci raných aminokyselin pro tvorbu β -skládaného listu, i když pouze s rozdílem ~5 % v porovnání s 20F.^{58,59} Všechny tyto metodiky jsou zatíženy různou chybovostí a zmíněná data je třeba interpretovat s opatrností. Zmíněná CD spektra byla například měřena z poměrně nízké koncentrace vzorku, která je spíše hraniční pro tuto metodu. V době sepisování této práce jsou měřena spektra pro 20F a 10E knihovnu v kratší délce a bez thrombin rozpoznávané sekvence, která se podařila produkovat v zhruba desetinásobném množství.

Analýzou míry tvorby terciární struktury u sekvencí v daných knihovnách bylo zjištěno, že knihovna 20F více podléhá štěpení, tedy míra tvorby terciárních struktur je u této knihovny pravděpodobně nižší. Nenaštěpených zůstává přibližně 20 % sekvencí, což se příliš neliší od výsledků podobné analýzy, kterou publikoval Chiarabelli a jeho výzkumná skupina v roce 2006⁵⁴. Knihovna 10E podléhá štěpení v menší míře, tedy pravděpodobnost tvorby terciárních struktur je zde vyšší. Tento výsledek by naznačoval, že rané aminokyseliny by ve skutečnosti měly větší potenciál pro terciární uspořádání. Vzhledem k tomu, že k raným aminokyselinám nepatří žádné aromatické aminokyseliny, které často upevňují hydrofobní jádro bílkovin, je toto pozorování neintuitivní. Nicméně mezi ranými aminokyselinami je řada hydrofobních aminokyselin, a navíc případné terciární uspořádání nemusí znamenat nutně pevné hydrofobní jádro. K podobnému závěru týkajícího se raných aminokyselin a tvorby sekundární a terciární struktury dokonce došla nedávná výpočetní studie A. D. Solise⁵⁰. Metoda proteolytického štěpení ovšem slouží spíše pro prvotní odhad tvorby terciárního uspořádání. Při ověřování její specifity se navíc zjistilo, že při vysokých koncentracích daného enzymu se objevuje nespecifické štěpení a množství použitého enzymu je pro proteolytické štěpení knihoven vzhledem k jejich koncentraci vyšší než množství použité pro štěpení kontrol specifity. I když je předpokládána tendence tvořit terciární strukturu u proteinových sekvencí knihoven vyšší než u kompletně nestrukturovaných kontrol specifity štěpení, nelze říci s jistotou, že jde o specifické štěpení daným enzymem. Vzhledem k tomu, že ani u knihoven ani u kontrol specifity štěpení nebyl fragment proteolýzy detekován metodou Western blot, a pro vizualizaci proteolýzy metodou SDS-PAGE gel nedisponujeme u knihoven dostatkem vzorku, bude v budoucnu třeba další ověření specifity použité metody. Navrhovaným dalším postupem je připravit knihovny stejným způsobem jak bylo popsáno v této práci, jen s výměnou specifického štěpného místa pro enzym thrombin za jinou definovanou sekvenci, kterou však tento enzym nebude schopen štěpit, podobně jako byly navrženy kontroly specifity štěpení. Tyto nové knihovny potom mohou sloužit jako negativní kontrola k posouzení míry specifity štěpení.

Závěrem byly v této práci úspěšně připraveny dvě knihovny náhodných sekvencí o různém aminokyselinovém složení. Knihovny byly úspěšně purifikovány. Jejich analýzou bylo zjištěno, že proteiny z raných aminokyselin mají zřejmě mírně vyšší preferenci pro vznik β -skládaného listu a zároveň také zřejmě větší potenciál pro tvorbu terciárního uspořádání. Výsledky prezentované

v této práci jsou součástí širšího projektu a tato práce rovněž navrhuje další postup optimalizace a ověření výsledků.

6. Závěr

Podařilo se připravit a purifikovat dvě kombinatoriální proteinové knihovny s různým aminokyselinovým složením. Obě knihovny se podařilo charakterizovat hmotnostní spektroskopií metodou MALDI-TOF a naměřit spektra cirkulárního dichroismu. Ze spekter cirkulárního dichroismu bylo možné částečně komentovat tendence k tvorbě sekundární struktury u sekvencí v obou knihovnách. Specifickým proteolytickým štěpením se nepodařilo dostatečně prokázat tendence proteinových sekvencí k tvorbě terciárních struktur, byl však navržen další postup pro vylepšení popisované metody.

7. Seznam literatury

1. Freeland, S. 'Terrestrial' Amino Acids and their Evolution. *Amino Acids, Peptides and Proteins in Organic Chemistry*. (2010). ISBN 9783527320967
2. Luisi, P. L. *The emergence of life: From chemical origins to synthetic biology. The Emergence of Life: From Chemical Origins to Synthetic Biology*. (2006). ISBN 9780511817540.
3. Meier, S. *et al.* Continuous Molecular Evolution of Protein-Domain Structures by Single Amino Acid Changes. *Curr. Biol.* **17**, 173–178 (2007).
4. Cech, T. R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **4**, 1–5 (2012).
5. Alva, V., Söding, J. & Lupas, A. N. A vocabulary of ancient peptides at the origin of folded proteins. *Elife* **4**, 1–19 (2015).
6. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
7. James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution - A 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368 (2003).
8. Holliday, G. L., Fischer, J. D., Mitchell, J. B. O. & Thornton, J. M. Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. *FEBS J.* **278**, 3835–3845 (2011).
9. Fox, G. E. Origin and evolution of the ribosome. *Cold Spring Harbor perspectives in biology* **2**, 1–19 (2010).
10. Kovacs, N. A., Petrov, A. S., Lanier, K. A. & Williams, L. D. Frozen in Time: The History of Proteins. *Mol. Biol. Evol.* **34**, 1252–1260 (2017).
11. Belousoff, M. J. *et al.* Ancient machinery embedded in the contemporary ribosome. *Biochemical Society Transactions* **38**, 422–427 (2010).
12. Hsiao, C., Mohan, S., Kalahar, B. K. & Williams, L. D. Peeling the onion: Ribosomes are ancient molecular fossils. *Mol. Biol. Evol.* **26**, 2415–2425 (2009).
13. Lupas, A. N. & Alva, V. Ribosomal proteins as documents of the transition from unstructured (poly)peptides to folded proteins. *J. Struct. Biol.* **198**, 74–81 (2017).
14. Burton, A. S., Stern, J. C., Elsila, J. E., Glavin, D. P. & Dworkin, J. P. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* **41**, 5459–5472 (2012).
15. Johnson, A. P. *et al.* The Miller volcanic spark discharge experiment. *Science* **322**, 404 (2008).
16. Bernstein, M. Prebiotic materials from on and off the early Earth. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1689–1700 (2006).
17. Miller, S. L., William Schopf, J. & Lazcano, A. Oparin's 'origin of life': Sixty years later.

- J. Mol. Evol.* **44**, 351–353 (1997).
18. Miller, S. L. The mechanism of synthesis of amino acids by electric discharges. *BBA - Biochim. Biophys. Acta* **23**, 480–489 (1957).
 19. Miller, S. L. & Urey, H. C. Organic Compound Synthesis on the Primitive Earth
Published by : American Association for the Advancement of Science. *Science* **130**, 245–251 (1959).
 20. Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H. J. & Miller, S. L. Prebiotic synthesis from CO atmospheres: Implications for the origins of life. *Proc. Natl. Acad. Sci.* **99**, 14628–14631 (2002).
 21. Yoshino, D., Hayatsu, K. & Anders, E. Origin of organic matter in early solar system-III. Amino acids: Catalytic synthesis. *Geochim. Cosmochim. Acta* **35**, 927–938 (1971).
 22. Bar-Nun, A., Bar-Nun, N., Bauer, S. H. & Sagan, C. Shock Synthesis of Amino Acids in Simulated Primitive Environments. *Science* **168**, 4–6 (1970).
 23. Lowe, C. U., Rees, M. W. & Markham, R. Synthesis of Complex Organic Compounds From Simple Precursor. *Nature* **199**, 219–222 (1963).
 24. Cleaves, H. J. The origin of the biologically coded amino acids. *J. Theor. Biol.* **263**, 490–498 (2010).
 25. Baross, J. & Hoffman, S. Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig. Life* **15**, 327–345 (1985).
 26. Amend, J. P. & Shock, E. L. Energetics of amino acid synthesis in hydrothermal ecosystems. *Science* **281**, 1659–1662 (1998).
 27. McMullin, E. R., Bergquist, D. C. & Fisher, C. R. Metazoans in Extreme Environments: Adaptations of Hydrothermal Vent and Hydrocarbon Seep Fauna. *Gravitational Sp. Biol. Bull.* **13**, 13–23 (2000).
 28. Ruhl, H. *et al. Delving Deeper Critical challenges for 21 st century.* (2015). ISBN 9789492043115.
 29. Marshall, W. L. Hydrothermal synthesis of Amino Acids. *Geochim. Cosmochim. Acta* **58**, 2099–2106 (1994).
 30. Faure, M., Lalevée, F., Gusokujima, Y., Iiyama, J.-T. & Cadet, J.-P. The pre-Cretaceous deep-seated tectonics of the Abukuma massif and its place in the structural framework of Japan. *Earth Planet. Sci. Lett.* **77**, 384–398 (1986).
 31. Chakrabarti, S. & Chakrabarti, S. K. Can DNA bases be produced during molecular cloud collapse? *Astron. Astrophysics* **354**, (2000).
 32. Zaia, D. A. M., Zaia, C. T. B. V & De Santana, H. Which amino acids should be used in prebiotic chemistry studies? *Orig. Life Evol. Biosph.* **38**, 469–488 (2008).
 33. Engel, M. H. & Nagy, B. Distribution and enantiomeric composition of amino acids in the Murchison Meteorite. *Nature* **269**, 837–840 (1982).

34. Shimoyama, A. & Ponnamperna, C. Amino Acids in the Yamato carbonaceous chondrite from Antarctica. *Nature* **282**, 394–396 (1979).
35. Muñoz Caro, G. M. *et al.* Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **416**, 403–406 (2002).
36. McGeoch, M. W., Dikler, S. & McGeoch, J. E. M. Hemolithin: a Meteoritic Protein containing Iron and Lithium. *Submitted* (2020).
37. Trifonov, E. N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **261**, 139–151 (2000).
38. Trifonov, E. N. The triplet code from first principles. *J. Biomol. Struct. Dyn.* **22**, 1–11 (2004).
39. Higgs, P. G. & Pudritz, R. E. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* **9**, 483–490 (2009).
40. Cronin, J. R. & Moore, C. B. Amino Acid Analyses of the Murchison, Murray, and Allende Carbonaceous Chondrites. *Science* **172**, 1327–1329 (1971).
41. Miller, S. L. & Orgel, L. E. *The Origins of Life on the Earth*. (1974). ISBN 9780136420828.
42. Lowe, C. U. Synthesis of Complex Organic Compounds From Simple Precursor. *Nature* **199**, 219–222(1963).
43. Fournier, G. P. & Alm, E. J. Ancestral Reconstruction of a Pre-LUCA Aminoacyl-tRNA Synthetase Ancestor Supports the Late Addition of Trp to the Genetic Code. *J. Mol. Evol.* **80**, 171–185 (2015).
44. Jordan, I. K. *et al.* A universal trend of amino acid gain and loss in protein evolution. *Nature* **435**, 528 (2005).
45. Granold, M., Hajieva, P., Toşa, M. I., Irimie, F. D. & Moosmann, B. Modern diversification of the amino acid repertoire driven by oxygen. *Proc. Natl. Acad. Sci.* **115**, 41–46 (2018).
46. Brack, A. & Orgel, L. E. β Structures of Alternating Polypeptides and Their Possible Prebiotic Significance. *Nature* **256**, 383–387 (1975).
47. Brack, A. & Spach, G. Multiconformational Synthetic Polypeptides. *J. Am. Chem. Soc.* **103**, 6319–6323 (1981).
48. Riddle, D. S. *et al.* Functional rapidly folding proteins from simplified amino acids sequences. *Nat. Struct. Biol.* **4**, 805–809 (1997).
49. Longo, L. M., Lee, J. & Blaber, M. Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. *Proc. Natl. Acad. Sci.* **110**, 2135–2139 (2013).
50. Solis, A. D. Reduced alphabet of prebiotic amino acids optimally encodes the conformational space of diverse extant protein folds. *BMC Evol. Biol.* **19**, 1–19 (2019).
51. Tanaka, J., Yanagawa, H. & Doi, N. Comparison of the frequency of functional SH3

- domains with different limited sets of amino acids using mRNA display. *PLoS One* **6**,1–8 (2011).
52. Newton, M. S., Morrone, D. J., Lee, K. H. & Seelig, B. Genetic Code Evolution Investigated through the Synthesis and Characterisation of Proteins from Reduced-Alphabet Libraries. *ChemBioChem* **20**, 846–856 (2019).
 53. Tanaka, J., Doi, N., Takashima, H. & Yanagawa, H. Comparative characterization of random-sequence proteins consisting of 5, 12, and 20 kinds of amino acids. *Protein Sci.* **19**, 786–795 (2010).
 54. De Lucrezia, D., Franchi, M., Chiarabelli, C., Gallori, E. & Luisi, P. L. Investigation of de novo totally random biosequences. Part IV. Folding properties of de novo, totally random RNAs. *Chem. Biodivers.* **3**, 869–877 (2006).
 55. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. B. T. Protein Structure Prediction Using Rosetta. *Numerical Computer Methods* **383**, 66–93 (Academic Press, 2004).
 56. Minervini, G. *et al.* Massive non-natural proteins structure prediction using grid technologies. *BMC Bioinformatics* **10**, 1–9 (2009).
 57. Prymula, K. *et al.* In silico structural study of random amino acid sequence proteins not present in nature. *Chem. Biodivers.* **6**, 2311–2336 (2009).
 58. Yu, J. F. *et al.* Natural protein sequences are more intrinsically disordered than random sequences. *Cell. Mol. Life Sci.* **73**, 2949–2957 (2016).
 59. Tretyachenko, V. *et al.* Random protein sequences can form defined secondary structures and are well-tolerated in vivo. *Sci. Rep.* **7**, 2–10 (2017).
 60. Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
 61. Neme, R., Amador, C., Yildirim, B., McConnell, E. & Tautz, D. Random sequences are an abundant source of bioactive RNAs or peptides. *Nat. Ecol. Evol.* **1**, 1–18 (2017).
 62. Knopp, M. & Andersson, D. I. No beneficial fitness effects of random peptides. *Nature Ecology and Evolution* **2**,1046–1047 (2018).
 63. Bornberg-Bauer, E. & Heames, B. Becoming a de novo gene. *Nat. Ecol. Evol.* **3**, 524–525 (2019).
 64. Davidson, A. R. & Sauer, R. T. Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl. Acad. Sci.* **91**, 2146–2150 (1994).
 65. Tretyachenko, V., Voráček, V., Souček, R., Fujishima, K. & Hlouchová, K. CoLiDe: Combinatorial Library Design tool for probing protein sequence space. *Submitted* (2020).
 66. Buck, M. Trifluoroethanol and colleagues: Cosolvents come of age. Recent studies with peptides and proteins. *Q. Rev. Biophys.* **31**, 297–355 (1998).
 67. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular

- dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260 (2000).
68. Provencher, S. W. & Glöckner, J. Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* **20**, 33–37 (1981).
 69. Ptáčková, B. Strukturní charakterizace vybraných náhodných proteinových sekvencí s vysokým obsahem neuspořádanosti. Diplomová práce (2018).
 70. Simpson, R. J. Zinc/Imidazole procedure for visualization of proteins in gels by negative staining. *CSH Protoc.* (2007).
 71. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119–139 (2005).

Dodatek

Sekvence oligonukleotidů použitých pro syntézu knihoven:

20F

>Vláknó 1

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGVWKVWKNB
GSHRVWKNNTDNCNBGDNCNBGVVAVVASHRVWKVWKVADNCVWKVVADNCDNCVVADNCDNC
NBGNNTDNCSHRVVANNTVWKDNCVVASHRVVANBGDNCGCGTTGGTTCCACGTGGCAGC

>Vláknó 2

TTTTTTTTTTCATTACTTATCGTCGTCATCTTTATAGTCANNTBBMWBCVNGNHANNTBBCVNGN
HTBBANNMWBGNHCVNYDSANNTBBTBBCVNMWBANNCVNTBBMWBMWBANNTBBMWBTBBYDS
GNHANNGNHANNTBBGNHMWBGNHMWBMWBYDSCVNTBBGNHYDSMWBGNHGCTGCCACGTGGA
ACCAACGC

10E

>Vláknó 1

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGMYYYYRGM
NRBYGMNYYRGNNYYRGMNYYRRBYRBYYYRGMNRBYGMNGMNMYYGNNRBYYYRYYRGNNRBY
MYYYRRBYGNNMYYYRRBYRBYGMNGMNGMNMYYGMNGCGTTGGTTCCACGTGGCAGCG

>Vláknó 2

TTACTTATCGTCGTCATCTTTATAGTCYRRRVYNKCNKCNKCNCRKRKVYRRKNNCYRRNKCR
VYYRRRRKRKNKCRVYYRRNKCRKRKRKVYRVYRVYYRRNKCNCRKNNCYRRNKCRVYNK
CYRRNKCRKNNCYRRRVYYRRNCCYRRYRRRRKYRRNCCGCTGCCACGTGGAACCAACGC

Sekvence kontrolních proteinů 665C a 665U:

>665U

```
ATGGCAACCAAAGGTGCAGATCATGGTCTGGCAGCACCGCAGCCGCATGCAAAATGGGATACCC
AGATTCCCTGCCGAAGGTGCCGATCGTGAACATCGTAGCGGTGGTGGTAATGAACGTCGCTTTTA
TAACGAAGGTGCAAAACATGCACAGGCAACCTGGGCAATTCCGGATGAACCGGCATTTTCATCTG
CAGCCTGCAGTTGGTGAAGGTGCCACCACCGATCAGGCAGGTAGCCTGGAAGATCAGTGGGTTC
GTAGCCTGAATAATAACGATGTTGATCCGACCCAGGCCGATGAAACCCTCGAG
```

>665C

```
ATGGCAACCAAAGGTGCAGATCATGGTCTGGCAGCACCGCAGCCGCATGCAAAATGGGATACCC
AGATTCCCTGCCGAAGGTGCCGATCGTGAACATCGTAGCGGTGGTGGTAATGAACGTCGCTTTTA
TAACGAAGGTGCAAAACTGGTTCCGCGTGGTAGTGCAATTCCGGATGAACCGGCATTTTCATCTG
CAGCCTGCAGTTGGTGAAGGTGCCACCACCGATCAGGCAGGTAGCCTGGAAGATCAGTGGGTTC
GTAGCCTGAATAATAACGATGTTGATCCGACCCAGGCCGATGAAACCCTCGAG
```

Sekvence použitých primerů:

>Primer A

```
CTGTAATACGACTCACTATATAGGG
```

>Primer B

```
CTATTAGTGGTGGTGGTGGTGGTGCTTATCATCATCATCTTTATAATC
```

>Primer C

```
AAAAACATATGGCAACCAAAGGT
```

>Primer D

```
AAAAACTCGAGGGTTTCATCGGC
```